NBER WORKING PAPER SERIES

ADAPTIVE CONSUMPTION BEHAVIOR

Peter Howitt
Ömer Özak

Working Paper 15427
http://www.nber.org/papers/w15427

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2009

We would like to thank participants at Brown University's theory and macroeconomics seminars, participants at the Econometric Society's LAMES 2008 and NASM 2009 meetings, and especially Kfir Eliaz, Glenn Loury, and David Weil for helpful comments and discussion. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Adaptive Consumption Behavior
Peter Howitt and Ömer Özak
NBER Working Paper No. 15427
October 2009
JEL No. E21,C63

ABSTRACT

This paper proposes and studies a theory of adaptive consumption behavior under income uncertainty and liquidity constraints. We assume that consumption is governed by a linear function of wealth, whose coefficients are revised each period by a procedure, which, although sophisticated, places few informational or computational demands on the consumer. We show that under a variety of settings, our procedure converges quickly to a set of coefficients with low welfare cost relative to a fully optimal nonlinear consumption function.

Peter Howitt
Department of Economics
Brown University, Box B
Providence, RI 02912
and NBER
peter_howitt@brown.edu

Ömer Özak
Department of Economics
Brown University, Box B
Providence, RI 02912
ozak@brown.edu

1 Introduction

The standard theory of lifetime utility maximization under uncertainty and liquidity constraints places enormous informational and computational demands on the consumer. Carroll (2001), a leading researcher in the area, presents this theory and argues that "when there is uncertainty about the future level of labor income, it appears to be impossible under plausible assumptions about the utility function to derive an explicit solution for consumption as a direct (analytical) function of the model's parameters". Similarly, Allen and Carroll (2001) admit that "finding the exact nonlinear consumption policy rule (as economists have done) is an extraordinarily difficult mathematical problem". This problem raises two closely related questions. First, are there simpler rules that have low welfare costs? And, second, can consumers learn the optimal rule or a simple low-cost rule? The first question has been studied with positive results for certain models (see Akerlof and Yellen, 1985a,b; Allen and Carroll, 2001; Cochrane, 1989). On the other hand, the second question has only been addressed, as far as we know, by Lettau and Uhlig (1999) and Allen and Carroll (2001), but with negative results.

In this paper, we propose an alternative adaptive theory of consumption behavior, which, in the spirit of Simon (1990), Arthur (1994) and Clark (1997), places limited demands on the consumer. We show that consumers who use this adaptive mechanism are able to learn a rule with a low welfare cost after a few periods. In particular, we show that constant-relative-risk-aversion consumers who follow a linear consumption rule in wealth and use our proposed algorithm, which adaptively adjusts the parameters of their rule, lose less than 0.5% of the equivalent consumption of the fully rational consumption rule within 500 periods with a probability higher than 0.9. Furthermore, we show that under social learning, the time required to attain a loss of 0.5% falls to less than 100 periods for some parametrizations. Additionally, the mean and median welfare losses, under both individual and social learning, fall to around 1% in less than 25 periods.[1]

[1] In an independent paper, Evans and McGough (2009) also address the question of adaptive approximation to optimal intertemporal choice. They present a procedure for updating expectations that is asymptotically fully optimal in a linear-quadratic environment, given that the decision maker knows enough about her environment to specify the correct functional form of her policy function. By contrast, we have found an adaptive procedure for updating the parameters of the policy function that works "reasonably well" even outside a linear-quadratic

The main problem with adaptive intertemporal choice is that there is no obvious simple criterion by which to measure the success of a rule one is trying out. Clearly, one has to look beyond the immediate utility flow it is generating, because the rule also has the potential of generating consumption in the future. But how does one evaluate that future potential in the same terms as the current utility flow, without undertaking computations that are as elaborate as those involved in solving the dynamic programming problem directly?

In both Lettau and Uhlig (1999) and Allen and Carroll (2001) the criterion of success is an estimate, based on past performance, of the discounted infinite sum of utilities. Lettau and Uhlig use a variant of Holland's (1992) classifier system, in which they call the measure of a rule's success its "strength". In each period after a rule has been used, its strength is adjusted partially towards the sum of the immediate utility attained under the rule last period plus the discounted strength of the rule that has succeeded it. They show however that the classifier system does a poor job of approximating optimal consumption behavior, even when the optimal consumption rule is available to the consumer. Their result follows from the fact that the system exhibits "state bias"; that is, it favors rules that apply in good states. Allen and Carroll, on the other hand, assume that the consumer is able to perform what amounts to a Monte Carlo simulation to evaluate each rule. They show that this procedure, instead of being quick and simple, actually needs 4 million periods in order to determine the optimal rule in their parameterization.[2] Accordingly, they argue that the procedure "is not an adequate description of the process by which consumers learn about consumer behavior" (Allen and Carroll, 2001, p.268).

Our approach follows Allen and Carroll in restricting consumers to rules that are linear in current wealth (for their parameterization they show that the optimal linear rule is almost as good as the optimal unrestricted rule). But instead of responding to a measure of cumulative discounted utility, we assume that the consumer adjusts her rule gradually in response to

environment when the decision-maker does not know the correct functional form. (Özak's (2009) analytical results show that our procedure will almost never be asymptotically fully optimal.)

[2] Allen and Carroll use a 5% threshold in order to assume the optimal rule has been found successfully. Under this assumption they require 1 million periods in order to get a success rate of 0.75 and of 4 million to get at least a success rate of 0.85.

the difference between the immediate marginal utility implied by the rule and the discounted marginal utility of next period's consumption. In effect our criterion of success is the ex post Euler equation error, and our algorithm operates like a stochastic approximation (see Robbins and Monro (1951), Ljung (1977) or Kushner and Yin (2003)) for solving the consumer's Euler equation.[3]

Although our approach presumes an awareness of sophisticated notions of Euler equations and numerical methods, nevertheless the informational and computational requirements of our algorithm are very low. Moreover, these requirements are independent of the size of the set of rules or states, which make it a good candidate for an adaptive procedure under bounded rationality for the problem at hand. In contrast, both requirements are increasing in the number of rules and states for Lettau and Uhlig and for Allen and Carroll. The reason for this difference is that in our algorithm, each period a consumer needs to revise the two parameters defining her linear rule, based only on its performance last period, whereas in the other papers, the consumer must to keep track of the past performance of a large number of rules under all states.

In our simulations, the algorithm converges with very high probability (at least 0.80, see table 4) in less than 500 periods for a set of parameters, which includes the same parameterization that took 4 million periods under the Carroll-Allen procedure. Although this is a big improvement vis-à-vis previous work, 500 periods is a long time for agents to learn. For example, Brown, Chua, and Camerer (2009) who study saving decisions in an experimental setting, find that their experimental subjects require around 120 periods to learn to behave optimally, with the average subject attaining around 80% of the optimal around that period. But in a world of boundedly rational consumers, clearly what matters to them in economic terms is not how long it takes to converge to a high degree of mathematical precision, but how long it takes before the welfare loss is small, say two or three percent. In those terms, even after 50 periods our mechanism imposes a small welfare cost on almost all consumers compared to the fully

[3]The idea that the opportunity cost of current spending could be learned adaptively through experience rather than calculated ex ante was suggested by Leijonhufvud (1993) in the context of Marshallian demand theory.

optimal nonlinear consumption function (see figures 12-21).

Even 50 years might seem too long for the procedure to be a reasonable description of opportunistic behavior on the part of the representative consumer. But in our view we should think of the representative consumer in statistical terms - as the mean or median consumer. This view accords with a long tradition in economics dating back at least to Marshall (1890, III,3.5), according to which economists cannot hope to account for the behavior of each individual with his or her idiosyncrasies, just the average behavior of large groups of individuals. In this respect it appears that our procedure approximates the optimal consumption behavior of a representative agent very well indeed; as we shall see in more detail below (tables 7-12), the mean or median welfare loss under our algorithm in most parameterizations is less than 2% after less than 10 periods.

In addition to our baseline simulations, we generalize our mechanism in order to study the effects of social learning, relaxation of credit constraints and changes to the agent's income process. Our brief analysis of social learning suggests that when imitation is allowed for, the time required for losses to be less than 0.5% with probability 0.9 falls to around 100 periods, while both the mean and median welfare losses become small even faster. On the other hand, we find that the relaxation of the credit constraint diminishes the incentives for the agents to learn a good rule and, thus, slows down their learning process. Interestingly, Brown, Chua, and Camerer (2009) find similar effects in their experiments.

Finally, we deal with the problem of changes in the agent's environment, in particular her income process. Obviously, in light of the Lucas critique, any adaptive mechanism must deal with the question of how to adapt to a change in the consumer's environment. We argue below that our procedure can be modified to make the consumer aware of changes in regime. Furthermore, since for some paramerizations she uses a constant gain adjustment procedure, even when she is not aware of the change, her ex post reaction to a regime change will be quick and in the appropriate direction.

The paper proceeds as follows: section 2 presents the model and the different consumption rules we use; section 3 presents the adaptive algorithm and its properties; section 4 presents the measures of welfare we use; section 5 shows the results of the simulations we conducted and section 6 concludes. All tables and figures are presented in appendix A.

2 The setting

The consumer's lifetime utility function is $U = \sum_{t=1}^{\infty} \delta^t u(c_t)$ where the period utility function is isoelastic, i.e.

$$u(c_t) = \begin{cases} \frac{c_t^{1-\theta}-1}{1-\theta} & \text{if } \theta \neq 1 \\ \ln c_t & \text{if } \theta = 1 \end{cases}$$

and the discount rate exceeds the rate of interest, so that

$$0 < \delta R < 1$$

where δ is the discount factor, $R \geq 1$ is the constant interest factor on one-period bonds. She starts each period t with wealth w_t, of which she consumes the amount c_t. She faces a liquidity constraint

$$c_t \leq w_t \tag{1}$$

and her wealth evolves according to the flow budget constraint

$$w_{t+1} = R\left(w_t - c_t\right) + y_{t+1},$$

where y_{t+1} is next period's income. We assume that income is an independently and identically distributed random variable with discrete support $\{y^i\}_{i=1}^n$ where $y^1 < y^2 < \ldots < y^n$ and the probability of each y^i is $p^i > 0$.

The consumer's behavior is determined by a consumption function

$$c_t = c\left(w_t\right),$$

which obeys the liquidity constraint (1). We assume that the consumption function is derived from a "notional" consumption function, $\widehat{c}\,(w_t)$, that ignores the liquidity constraint, so that

$$c\,(w_t) = \min\left\{\widehat{c}\,(w_t)\,, w_t\right\}.$$

We refer to the notional function $\widehat{c}(\cdot)$ as the consumer's "rule". In what follows, we assume that $\widehat{c}(\cdot)$ is increasing and concave, and satisfies

Assumption A. *There exists $\widetilde{w} > 0$ such that $R\,(w - \widehat{c}(w)) + y^n < w$ for all $w > \widetilde{w}$.*

This last assumption guarantees that the consumer's wealth will be bounded above by $\widetilde{w}$ in the long run. Specifically, theorem 1.2 in Özak (2009, p.7) assures that under the above assumptions there is a unique invariant wealth distribution, π, whose support is contained in the interval $[y^1, \widetilde{w}]$.

Additionally, we assume that

$$\widehat{c}\,(0) \geq 0 \text{ and} \tag{2}$$

$$\widehat{c}\,(\overline{w}) - \overline{w} = 0 \text{ for some } \overline{w} \in \left(y^1, y^n\right) \tag{3}$$

This implies that when wealth surpasses $\overline{w}$ the consumer is no longer liquidity constrained, so that

$$c\,(w) = \begin{cases} w & \text{if } w \leq \overline{w} \\ \widehat{c}\,(w) < w & \text{if } w > \overline{w} \end{cases}.$$

We refer to $\overline{w}$ as the consumer's "crossover wealth". The assumption that $\overline{w} > y^1$ requires the liquidity constraint to be binding for at least some observable wealth levels in the long-run, for otherwise, although the agent might start off at a wealth level for which she is constrained, she would immediately after the first period get income and wealth levels under which she could never again be constrained.[4] On the other hand, the assumption that $\overline{w} < y^n$ requires the consumer to save with positive probability in the long run, since otherwise (given Assumption

[4]This follows directly from the liquidity constraint (1) and the flow budget constraint, which together imply that w_t, $t \geq 1$, cannot fall below y^1.

A) she would in finite time end up with $c_t = w_t = y_t$ for all t. We call a rule $\widehat{c}(\cdot)$ *admissible* if it is increasing, concave and satisfies (2), (3) and Assumption A. The adaptive algorithm we specify below for revising the consumer's rule ensures that the rule always remains admissible. In the next two subsections we show conditions under which the fully optimal consumption rule and a linear rule are admissible.

2.1 Optimal consumption

The optimal consumption function $c^*(w)$ can be derived from the dynamic programming problem:

$$V(w) = \max_{c \leq w} \{u(c) + \delta E_y V(R(w-c)+y)\}$$

where E_y is the expectation with respect to income y. This corresponds to the notional function

$$\widehat{c}^*(w) = \arg\max_{c} \{u(c) + \delta E_y V(R(w-c)+y)\},$$

which we refer to as the "fully optimal" rule. For future reference, note that the first-order condition defining $\widehat{c}^*(w)$ is

$$u'(c) = E_y q,$$

where q is the marginal continuation value

$$q = \delta R u'\Big(c^*\big(R(w-c)+y\big)\Big)$$

whose value is not known when c is chosen because it depends on next period's income y.

It is known that under our assumptions the fully optimal rule $\widehat{c}^*(\cdot)$ is indeed increasing and concave, that it satisfies (2) and Assumption A, and that for δR large enough (i.e. if the consumer is patient enough or the interest rate is high enough) it also satisfies (3) and hence is admissible (see e.g. Carroll, 2004; Carroll and Kimball, 1996; Özak, 2009). From now on we assume that the consumer is indeed patient enough, so that $\widehat{c}^*(\cdot)$ satisfies (3) and is thus admissible.

2.2 Linear consumption

We follow Carroll and Allen (2001) in assuming that the consumer's consumption rule is linear with coefficients $\gamma = (\alpha, \beta)$, so that

$$\widehat{c}^{\gamma}(w) = \alpha + \beta w.$$

In order for $\widehat{c}^{\gamma}(\cdot)$ to be increasing and to obey (2) we need α and β both to be non-negative. In order to satisfy (3) we also need the marginal propensity to consume β to be less than unity. The consumer's crossover wealth is then

$$\overline{w}^{\gamma} = \frac{\alpha}{1-\beta} > 0$$

and the consumption function can be written as

$$c^{\gamma}(w) = \min\left\{\overline{w}^{\gamma} + \beta\left(w - \overline{w}^{\gamma}\right), w\right\}$$

Assumption (3) also requires

$$(1-\beta)\, y^{1} < \alpha < (1-\beta)\, y^{n} \tag{4}$$

and Assumption A requires

$$\beta > \frac{R-1}{R} \tag{5}$$

We say that a coefficient vector γ is admissible if the rule $\widehat{c}^{\gamma}(\cdot)$ is admissible. Thus, we see that

Proposition 1. *The coefficients (α, β) are admissible if and only if $\alpha \geq 0$, $\beta < 1$ and they satisfy (4) and (5).*

The set of admissible coefficients is denoted $\mathcal{A}$ and is illustrated by the shaded triangular area in Figure 1 below.

3 Revising the coefficients

The consumer chooses c_t according to the consumption function defined by the linear rule $\widehat{c}^{\gamma_t}(\cdot)$ with coefficients $\gamma_t = (\alpha_t, \beta_t)$, i.e.

$$c_t = \min\{\alpha_t + \beta_t w_t, w_t\}$$

Then income y_{t+1} is realized and next period's wealth becomes

$$w_{t+1} = R(w_t - c_t) + y_{t+1}$$

Before choosing next period's consumption the consumer has an opportunity to revise the coefficients γ_t. In doing so she first computes the "error" she made in period $t-1$

$$e_{t-1} = q_t - u'(\alpha_{t-1} + \beta_{t-1} w_{t-1})$$

where q_t is the realized marginal continuation value

$$q_t = \delta R u'(c_t)$$

Clearly if $e_{t-1} = 0$ then if she had known that the marginal continuation value was going to be q_t she would have been happy with her choice of coefficients, because it led her to choose a notional consumption whose marginal utility was just equal to its marginal cost. In this sense, even with hindsight she did not make a mistake. On the other hand if $e_{t-1} \neq 0$ she made an ex post error of consuming too much (if $e_{t-1} > 0$) or too little (if $e_{t-1} < 0$).

Accordingly, we suppose that her revisions will depend on this error according to the following multi-step procedure. She begins with a symmetric 2×2 "moment" matrix M_{t-1}, and then goes through the following steps:

Algorithm.

1. *Choose a new moment matrix M_t using the formula*

$$M_t = (1-\varepsilon) M_{t-1} + \left[\xi e_{t-1} u''' (\alpha_{t-1} + \beta_{t-1} w_{t-1}) + u'' (\alpha_{t-1} + \beta_{t-1} w_{t-1})^2\right] \begin{pmatrix} 1 & w_{t-1} \\ w_{t-1} & w_{t-1}^2 \end{pmatrix}$$

where $\varepsilon \in [0,1)$ is a constant gain parameter, $\xi \in \{0,1\}$ is a parameter allowing further simplification of the procedure.[5]

2. *If M_t is well conditioned*[6] *choose a provisional coefficient vector $\gamma^p_{t+1} = \left(\alpha^p_{t+1}, \beta^p_{t+1}\right)$ according to*

$$\begin{pmatrix} \alpha^p_{t+1} \\ \beta^p_{t+1} \end{pmatrix} = \begin{pmatrix} \alpha^p_t \\ \beta^p_t \end{pmatrix} + M_t^{-1} \left[e_{t-1} u'' (\alpha_{t-1} + \beta_{t-1} w_{t-1})\right] \begin{pmatrix} 1 \\ w_{t-1} \end{pmatrix}$$

3. *If M_t is not well conditioned or if the provisional γ^p_{t+1} chosen in step 1 is not admissible, choose the nearest vector to γ_t that would have eliminated the most recent error, i.e. let*

$$\gamma^p_{t+1} = \arg\min_\gamma (\gamma - \gamma_t)^2 \text{ \textit{subj to} } q_t - u' (\alpha + \beta w_{t-1}) = 0$$

4. *If γ^p_{t+1} is still inadmissible, set $\gamma^p_{t+1} = \gamma_t$.*

5. *Shrink the step size by a factor $\eta \in (0,1]$ and set the new coefficients according to*

$$\gamma_{t+1} = \gamma_t + \eta \left(\gamma^p_{t+1} - \gamma_t\right).$$

[5] If $\varepsilon = 0$ this algorithm can be written as a decreasing gain algorithm. If $\xi = 0$ then what would have been a quasi-Newton method, as explained in the next section, becomes what is known in numerical analysis as a quasi-Gauss-Newton method, which obviates the need for calculating the third derivative u'''.

[6] Specifically, if the condition number $r2(M_t)$ is less than the conventional limit 10^{10}, indicating that the matrix is reliably nonsingular. See Judd (1998).

3.1 Numerical motivation

The first two parts of this procedure (assuming $\xi = 1$) constitute a simplified, recursive variant of the standard Newton method for solving the nonlinear weighted least squares problem of choosing (α, β) to minimize the weighted sum of squared errors:

$$\frac{1}{t}\sum_{k=1}^{k=t} \lambda^{t-k} \left(u'\left(\alpha + \beta w_{k-1}\right) - q_k\right)^2$$

where $\lambda = 1-\varepsilon$, given the history of marginal continuation values and wealth levels $\{w_{k-1}, q_k\}_{k=1}^{t}$. The first iteration in Newton's method starting at $\alpha_{t-1}, \beta_{t-1}$ would be

$$\begin{pmatrix} \Delta\alpha_t \\ \Delta\beta_t \end{pmatrix} = H_t^{-1} \sum_{k=1}^{k=t} \lambda^{t-k} \left[e_{k-1} u''\left(\alpha_{t-1} + \beta_{t-1} w_{k-1}\right)\right] \begin{pmatrix} 1 \\ w_{k-1} \end{pmatrix}$$

where H_t is the Hessian matrix

$$H_t = \sum_{k=1}^{k=t} \lambda^{t-k} \left[e_{k-1} u'''\left(\alpha_{t-1} + \beta_{t-1} w_{k-1}\right) + u''\left(\alpha_{t-1} + \beta_{t-1} w_{k-1}\right)^2\right] \begin{pmatrix} 1 & w_{k-1} \\ w_{k-1} & w_{k-1}^2 \end{pmatrix}$$

Step 2 above produces the same iteration except that (a) it uses only the term involving the most recent error e_{t-1} whereas the Newton iteration considers all past errors $\{e_{k-1}\}_1^{t-1}$, and (b) it uses the moment matrix M_t rather than the Hessian H_t. Moreover, the formula in step 1 above for changing M_t is the same as the formula that describes the change in H_t when observation $t-1$ is added. Thus the first two steps are equivalent to taking a single Newton iteration but failing to use information from past errors and then failing to update the parameter values $(\alpha_{k-1}, \beta_{k-1})$ in any but the most recent term of the Hessian.

The fallback procedure in step 3 is a myopic steepest-descent method for solving the same nonlinear least-squares problem. It works as a provisional "projection facility" in case the recursive quasi-Newton method, which depends on past history as embodied in the moment matrix fails; specifically, in that case she chooses to ignore history and minimize the most recent error. This is in the spirit of the Levenberg-Marquard method for solving systems of equations, which combines the Newton method with the steepest-descent method in each step. Here the steepest descent method is used only as a fallback option.

In step 4 the consumer gives up if she still cannot find an admissible coefficient vector. Step 5 is a commonly used prudential measure to reduce the danger of instability from overreacting to new information.

3.1.1 Q learning

The procedure is also related to what is known in the Artificial Intelligence literature as q learning (Watkins, 1989). In the present context, q learning would have the consumer attempt to estimate a relationship between w_{t-1} and q_t, perhaps through least squares learning or neural networks. Each period she would choose c_t according to the rule

$$u'(c_t) = \widehat{E}_t(q_t|w_{t-1})$$

where $\widehat{E}_t(q_t|w_{t-1})$ is the estimated value of q_t given the current wealth level w_{t-1} according to her most recently estimated relationship.

3.2 Informational and computational requirements

The procedure outlined above requires a certain amount of sophistication, in the sense that moment matrices, Newton methods, condition numbers, Euler equations are not familiar household names. The consumer must also be sophisticated enough to realize that (a) if her crossover wealth is less than y^1 then her liquidity constraint will never bind, (b) if her marginal propensity to consume is less than $(R-1)/R$ then her wealth will accumulate with no upper limit and (c) if her crossover wealth is greater than y^n then she will eventually reach a situation in which her liquidity constraint is binding from then on and will never save. Moreover she needs to realize that all three of these outcomes are suboptimal for someone with a rate of time preference greater than the rate of interest but small enough to warrant precautionary saving, and hence that she should restrict herself to admissible rules. Although this level of economic sophistication might seem excessive for a boundedly rational agent, we show below that dispensing with it, i.e. allowing for rules in a superset of the set of admissible rules, does not change our results dramatically.

Nevertheless the procedure makes relatively few informational or computational demands on the consumer, especially in comparison to the demands involved in calculating the optimal consumption function. This is an important consideration for intelligent behavior in a world where information storage capacity and computational time are scarce resources. In particular, all the consumer needs to know is her lifetime utility function, the interest factor R, the gain parameter ε, the shrinkage factor η and the minimal and maximal possible income levels y^1 and y^n. Each period she must remember only 10 numbers: the two most recent coefficient vectors γ_t and γ_{t-1}, the most recent marginal continuation value q_t, the two most recent wealth levels w_t and w_{t-1}, and the three elements defining the most recent moment matrix M_{t-1}. In addition to elementary addition and multiplication, she just needs to be able to compute the first three derivatives of her utility function, to determine the conditioning value of a 2×2 matrix and to compute its inverse.

4 Welfare cost

We use two different indices to measure the welfare cost for a consumer of following a specific linear consumption rule rather than the fully optimal rule. Both indices are based on ex-ante equivalent consumption for an agent following a certain rule, starting from a randomly assigned wealth; but each index uses a different probability distribution for assigning initial wealth. The first index uses the stationary distribution implied by the fully optimal consumption rule, whereas the second one uses the stationary distribution implied by the specific linear rule. In either case, the index measures the percentage difference in certainty equivalent consumption between the fully optimal rule and the linear rule. As will be seen, the two indices produce very similar results.

More specifically, suppose that initial wealth w_0 is assigned randomly according to some distribution λ. The ex ante expected lifetime utility of a consumer using the fully optimal rule is given by

$$EV^* \equiv \int_W V(w)\lambda(dw) \tag{EV^*}$$

where the value function V is defined in section xxx above. Thus we can define the certainty equivalent consumption of the fully optimal rule as

$$CE^* \equiv u^{-1}(EV^* \cdot (1-\delta)) = \Big[1 + (1-\theta)(1-\delta)EV^*\Big]^{\frac{1}{1-\theta}} \tag{CE^*}$$

For any given w_0, the expected life-time utility of a consumer using the specific linear rule $\widehat{c}^{\gamma}(w)$ with parameters $\gamma = (\alpha, \beta)$, is given by

$$U^{\gamma}(w_0) = \sum_{t=0}^{\infty} E_0\Big[\delta^t u\Big(\min\{\alpha + \beta w_t, w_t\}\Big)\Big], \tag{6}$$

where w_t evolves according to the flow budget constraint. So, given the distribution λ of w_0, the ex-ante expected lifetime utility and the certainty equivalent consumption for this specific rule are given by

$$EV^{\gamma} \equiv \int_W U^{\gamma}(w)\lambda(dw) \text{ and} \tag{EV^{γ}}$$

$$CE^{\gamma} \equiv u^{-1}(EV^{\gamma} \cdot (1-\delta)) = \Big[1 + (1-\theta)(1-\delta)EV^{\gamma}\Big]^{\frac{1}{1-\theta}} \tag{CE^{γ}}$$

If the wealth process generated by $c^*(w)$ satisfies Assumption A, then there exists a unique ergodic invariant distribution over wealth π_*. In this case, let EV_*^*, CE_*^*, EV_*^{γ} and CE_*^{γ} be the values implied respectively by (EV^*), (CE^*), (EV^{γ}) and (CE^{γ}) when $\lambda = \pi_*$. Thus our first index of welfare cost for a consumer using the rule $\widehat{c}^{\gamma}$ is

$$D_1^{\gamma} = \frac{EC_*^* - EC_*^{\gamma}}{EC_*^*} * 100.$$

If the linear consumption rule satisfies Assumption A, let π_{γ} be the unique invariant distribution determined by the rule and let EV_{γ}^*, CE_{γ}^*, EV_{γ}^{γ} and CE_{γ}^{γ} be the values when $\lambda = \pi_{\gamma}$. Thus our second index of welfare cost for a consumer using the rule $\widehat{c}^{\gamma}$ is

$$D_2^{\gamma} = \frac{EC_{\gamma}^{*} - EC_{\gamma}^{\gamma}}{EC_{\gamma}^{*}} * 100.$$

5 Numerical results

5.1 Baseline scenario

In order to study the behavior of the algorithm, we simulate the model for a set of values of the CRRA parameter θ and the discount factor δ, using the same interest factor R and the same income process but with different initial conditions. We used the income process studied by Allen and Carroll (2001), which according to these authors "matches (very roughly) the empirical evidence on the amount of transitory variation in annual household income observed in the *Panel Study of Income Dynamics*". The income process is defined by $(y^1, y^2, y^3) = (0.7, 1.0, 1.3)$ with probabilities $(p^1, p^2, p^3) = (0.2, 0.6, 0.2)$ respectively. We also assumed $R = 1$ as these authors do, allowing us to compare our algorithm with theirs in a transparent way. We took $\delta \in \{0.9, 0.95\}$, $\theta \in \{1.5, 2, 3.0, 3.5, 4\}$, which includes the values Allen and Carroll (2001) assumed in their work ($\delta = 0.95$ and $\theta = 3$), $\eta \in \{0.5, 1\}$, $\xi \in \{0, 1\}$ and $\varepsilon \in \{0, 0.2\}$. We evaluated the linear rules in the $[0, 2] \times [0, 2]$ space with a grid of 40,000 points, each separated at a distance of 0.01. Finally, we analyzed all rules, both the optimal one and the linear ones, constraining wealth to be in $[0, 5]$ at all times.

5.1.1 Consumption rules and welfare

As a first step, we calculated the optimal consumption function $c^*(w)$ for each parameter configuration. Figure 2 shows the rational consumption function c^* for each parameter configuration. Under our assumptions, we have that Assumption A is satisfied if $c^*(w) = 1.3$ for some w, which as the figure shows, clearly holds for all our parameters, so there exists a unique distribution π_* for each parameter configuration, allowing us to calculate EV_*^* and CE_*^*. As can be seen in table 1, EV_*^* is decreasing in both δ and θ, while CE_*^* is decreasing in θ and increasing in δ.

In table 3 we present the optimal linear rule for each set of parameters. As can be seen there, all optimal rules imply that Assumption A holds, so we can also compute EV_*^γ, EC_*^γ, EV_γ^γ, EC_γ^γ, EV_γ^* and CE_γ^*. Our calculations show that the behavior of all EV's and EC's is similar to EV_*^* and EC_*^* with respect to the underlying parameters, i.e. all EV's decrease in both θ and δ, while all EC's decrease in θ and increase in δ.

As can be seen in figures 3-7, most linear rules in $[0,2] \times [0,2]$ generally have a low cost, according to both our indices. Furthermore, as figures 8-11 show, the set of consumption rules that achieve a percentage deviation less than or equal to 0.5% under both the optimal and the actual distribution of initial wealth is compact and of positive Lebesgue measure. The optimal linear rule belongs to this set, has a marginal propensity to consume in the range 0.2-0.4 and its costs are in the range 0.2%-0.3% for our set of parameters (see table 3), which is very low and generalizes the case studied by Allen and Carroll (2001). Table 3 also shows the loss incurred when the agent follows the "consume everything" rule $\gamma = (0, 1)$ and the expected loss of taking a rule at random from the whole set of parameters and from the admissible set.[7] As can be seen there, if the agent figures out that she should always have a rule in the admissible set, she can lower her expected loss by almost 40%. On the other hand, if she follows the "consume everything" rule, which is the simplest of all rules, she can lower her loss even below her expected loss in the admissible set. This is an interesting result, which we are not aware has been pointed out in the literature, and which might be an explanation as to why people save less than optimally. If consumers are boundedly rational and try a random rule in order to see if it performs better than the consume everything rule, they will be disappointed in general, which will favor the continued use of the consume everything rule. This feature makes it a good convention to follow. Still, there are big gains to be attained from figuring out the optimal consumption rule, which can lower losses three to twentyfold compared to the consume everything rule.

[7] Cochrane (1989) realizes similar calculations and finds even smaller levels of losses. This is due in part to the fact that he uses aggregate income in order to simulate the income process of the agent. As will be seen below, the support of the distribution of the income process has a big impact on the welfare losses each linear rule generates.

5.1.2 Adaptive behavior

In order to understand the behavior of the algorithm we ran 20,000 simulations for each parameter configuration with random initial wealth and a random initial admissible linear rule. Figures 12-21 show the behavior of the distribution of welfare losses $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ across time for all simulations for $\varepsilon = 0$, $\xi = 1$ and $\eta = 1$, where $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ are calculated for the rule with parameters $\gamma_t = (\alpha_t, \beta_t)$. The behavior of the distribution is summarized in these figures by the maximum and minimum loss (black lines), the mean (blue), the median (green) and the 25-th, 75-th, 90-th, 95-th, 99-th and 99.9-th percentiles (red lines). Notice that the values of all these measures decrease for the first 100-250 periods and then follow a flatter trajectory, which seems to indicate that within that time frame the algorithm achieves its stationary distribution, which is more concentrated around the mean.

In those same figures (12-21) we present the behavior of the consumption rule parameters and their distribution. The story here is also similar, showing convergence towards a long-run distribution of the parameters and a higher concentration of the probability of those parameters around their mean in roughly 250 periods or less. In order to better appreciate what is happening, in table 4 we show for periods 0, 50, 100, 250 and 500, the probability of having a loss less than or equal to 0.5%, i.e. the probability that the consumption rule in that period belongs to the sets identified in figures 8-11. As can be seen in this table, the probability rises extremely fast, going from almost zero in period 0 for most parameter configurations, to values above 30-40% in period 50, increasing above 40-60% in period 100, reaching levels above 80% by period 250 and above 90% by period 500. Clearly there is much variation in the speed of convergence to these sets, with the cases $(\delta, \theta) = (0.9, 1.5)$ and $(\delta, \theta) = (0.95, 1.5)$ being the ones with the highest rates of convergence and the cases $(\delta, \theta) = (0.9, 4)$ and $(\delta, \theta) = (0.95, 3.5)$ with the slowest convergence rates. This seems to be explained by the effect that both increases in δ and in θ have on the size of the set of parameters that achieve a loss of less than 0.5%. However, it is not clear how changes in the underlying parameters affect the speed of convergence, though this relation seems highly non-linear. In part, this seems to be generated by the fact that changes in θ and δ have similar effects on the curvature of c^*, where increases in δ or in θ make the function more concave. So, two functions, one with a high value of δ and a low θ and the other with a low δ and high θ might be very similar on the domain of interest (see e.g. the rational consumption functions for $(\delta, \theta) = (0.9, 4)$ and $(\delta, \theta) = (0.95, 2)$, or, $(\delta, \theta) = (0.9, 3)$ and $(\delta, \theta) = (0.95, 1.5)$), so that their respective rates of convergence are also similar.

We repeated the simulations assuming different values of (ε, η, ξ). Given the overall similarity of the results we do not present them here in detail, but limit ourselves to highlight the major differences with our previous simulations.[8] Under this new set of parameters, the behavior of the distributions of losses and parameters in terms of convergence to a stationary distribution within 250 periods was similar as before, though the dispersion around the mean increased and the speed of convergence decreased especially for $\varepsilon = 0.2$. One striking effect of this was that the long-run probability of having a consumption rule in the set of consumption rules with losses less than or equal to 0.5% fell, in some cases dramatically, and stayed stationary at that level without any tendency to converge towards 1 as was the case before. Still, this probability was bounded away from zero for all the simulations we realized. In table 5 we compare the different trajectories for the case $\theta = 3.5$ and $\delta = 0.95$, which was the case with the slowest convergence rate in our baseline simulations.

5.1.3 Social Learning

We additionally looked at the effects of allowing agents to learn through social interaction. To do so, we set up agents on circles of 25, 50, 100, 200 individuals and allowed each agent to interact with his left, right or both left and right neighbors. Each agent was allowed to see her neighbor's γ^p_{t+1}. Let i denote the agent and N_i the set of neighbors of agent i, $\phi \in [0, 1]$, then step 5 of the algorithm works in the following way

(5') Shrink the step size by a factor $\eta \in (0, 1]$:

$$\gamma_{i,t+1} = \gamma_{i,t} + \eta\Big(\phi\gamma^p_{i,t+1} + \frac{(1-\phi)}{|N_i|}\sum_{j\in N_i}\gamma^p_{j,t+1} - \gamma_{i,t}\Big)$$

[8] These results can be obtained from the authors by request.

Allowing for this type of interaction speeds up and increases the probability of convergence of the algorithm significantly as can be seen in table 6. Since neighbor's initial conditions were randomly set, this type of social learning can be interpreted as a "keeping up with the Jones'" scheme, where each agent considers each of her neighbors as the Jones family.

5.1.4 Credit constraints

Additionally, we analyzed how the credit constraint affected learning. So, we relaxed the credit constraint allowing agents to consume at most $w_t + B$ every period, where we took $B \in \{0.1, 0.3\}$. These simulations showed that the more the constraint was relaxed the slower agents seemed to learn. This result can be seen as a confirmation of the conjecture proposed by Satz and Ferejohn (1994), who argue that the more constrained agents are, the more powerful their incentives to behave rationally. Overall we found that the behavior of this new set of simulations was similar to the ones described previously, i.e. the less rational agents were assumed to be, the slower they learned, though they *did learn* nonetheless.

5.2 Shocks to income

In all the previous simulations we held the income process fixed and changed the different parameters of the model. In this subsection we hold the parameters fixed at $\theta = 3.5$, $\delta = 0.95$, $B = 0$, $\eta = 1$, $\epsilon = 0$, $\xi = 1$ and $\phi = 1$, but change the income process. We analyze 4 additional income processes and compare the behavior of the algorithm and the implied welfare costs under these new processes with the one implied by the original process. In order to have comparable results across simulations, we now allow wealth to be in the range $[0, 10]$ and consider a grid of wealth levels 0.0025 apart. The income processes we consider are:

$$Y^1 = (0.7, 1, 1.3) \qquad P^1 = (0.2, 0.6, 0.2)$$

$$Y^2 = (1.4, 2, 2.6) \qquad P^2 = (0.2, 0.6, 0.2)$$

$$Y^3 = (1, 1.4, 2, 4.1) \qquad P^3 = (0.1, 0.2, 0.6, 0.1)$$

$$Y^4 = (0.3, 0.7, 1, 2.1) \qquad P^4 = (0.05, 0.25, 0.6, 0.1)$$

$$Y^5 = (0.1, 0.7, 1, 1.3, 1.9) \qquad P^5 = (0.05, 0.15, 0.6, 0.15, 0.05)$$

Figure 22 shows the optimal consumption rules for these cases, while table 13 presents the losses and optimal rules for these income processes. As can be seen there, the differences in the income processes generate quite big and striking differences in the expected losses an agent faces under different rules. For some income processes, e.g. Y^4 and Y^5, only very few linear rules have low expected losses, as can be seen in figures 23-26. This implies that using a random rule is quite costly. Also, the consume everything rule fares badly, since in these cases the lowest income is close to zero, generating huge losses for the agent if she ends up being wealth constrained. On the contrary, the optimal linear rule for each process still has a relatively low associated loss. Furthermore, the marginal propensity to consume stays in the same range across cases (ca. 0.17-0.25). These results imply that a simple adaptive algorithm, as the one proposed in this paper, can have big welfare effects for boundedly rational agents, if the algorithm converges to low welfare losses in general. Table 14 shows the evolution of the probability of having a rule with a loss lower than 3% when using the algorithm. We increased the range to 3% given that for some of the income processes no rule has a loss less than 0.5% and almost no rule has a loss less than 1%. As can be seen in table 14, the qualitative dynamics are very similar to the ones we had found in our original setup, so that the rate of convergence is similar to our baseline simulation for most processes. The exception is Y^5 for which convergence requires close to double the time as in the other scenarios.

Until now, we have assumed that agents do not use any other information of the income process except its lowest and highest possible levels. One way in which this might slow convergence to the optimal rule is that the agent cannot distinguish a situation where the Euler error e_{t-1} is high because of a bad income draw from a situation in which e_{t-1} is high because of a bad rule. Furthermore, the agent would not notice any changes in the stochastic process determining income, if these do not affect the range of values that can occur. This is especially important in the case where $\epsilon = 0$, since this would imply that if she had been learning for while, it would be difficult for her to change her behavior. On the other hand, if $\epsilon > 0$ the agent's behavior never settles down, so that she will incorporate any changes in her environment into her behavior, *even if she does not realize there has been a change.*

In order to deal with this issue, we now assume that agents update their parameters in the same way as before, but change the manner in which they calculate q_t. Instead of using the realized continuation value, an agent uses the average realization of the continuation value under their current parameters, i.e.

$$q_t = \delta R \sum_{i=1}^{n} p^i u'(c_t^i), \tag{ID}$$

where c_t^i is the amount she would have consumed if she had received income level y^i in period t, i.e.

$$c_t^i = \min\left\{\alpha_t + \beta_t(R(w_{t-1} - c_{t-1}) + y^i), R(w_{t-1} - c_{t-1}) + y^i\right\}, \quad i = 1, \dots, n.$$

This change requires the agent to know the whole distribution of the income process, so that $2(n-1)$ additional numbers have to be remembered by the agent. Table 15 shows the evolution of the probability of being at a loss lower than 3%. As expected, the time to convergence is generally lowered, or equivalently, the probability of being close to the optimal rule is increased for any period for all income processes studied.

This modification to our algorithm inoculates it against the Lucas critique. That is, the agent can now respond immediately to a change in regime by modifying her behavior accordingly. To see how this works, we analyzed the effects of a shock to income in which the income process changes to some other one in the set of processes we have studied, and then 25 years later returns unexpectedly to the original process, assuming that the consumer is informed immediately of each change in the income process. Given that we have assumed that agents are sophisticated enough to know that the optimal rule lies in the set $\mathcal{A}_Y$ of admissible rules, which depends on the income process, we need an assumption specifying the way in which agents react to the new information of a change in the income process. One possible assumption is that agents dismiss all their accumulated experience up to that point and start the process from scratch. This amounts to almost the same exercise we have done in the previous simulations, except that the initial distribution of wealth will be closer to the stationary distribution of the original process. Given the fact that initial wealth conditions do not affect the long-run evolution of the system, we do not need to analyze this scenario again. Instead, we assume that agents keep their marginal propensity to consume β, while changing α in such a way as to keep the relative position of the rule in the new set $\mathcal{A}_{Y'}$ similar to the one in $\mathcal{A}_Y$. In particular, assuming that the agent's rule is $\gamma_t = (\alpha_t, \beta_t)$ at the moment of the shock, she can determine the value $\psi \in (0,1)$ such that $\beta_t = \alpha_t * \Big(\psi y^1 + (1-\psi)y^n\Big)$, and then use that value in order to find her new starting $\alpha_t' = \frac{\beta_t}{\psi {y'}^1 + (1-\psi){y'}^n}$. Figures 27-30 show the dynamics when agents' income follows Y^1 in periods 1-25 and 51-100, while it follows Y^i, $i \neq 1$ during periods 26-50. As can be seen there, a permanent shock to income has various effects. First, the distribution of consumption, wealth, (α_t, β_t), and losses become more disperse during the shock. Second, welfare losses overshoot, and only slowly decrease towards the minimum loss. Interestingly, the effect on welfare is bigger on the average than on the median. This seems to come from the huge difference between the low losses implied by good rules and the big losses generated by bad rules. Thus, even the existence of one individual with a significantly bad loss, increases the average loss towards high values, even though most individuals might be having low losses. Third, although the effect on consumption comes mostly from the change in the value of the intercept, which jumps discretely at the moment of the shock, it is the distribution of the marginal propensity to consume that changes the most during the shock period.

6 Conclusion

The assumption of complete rationality has increasingly been criticized due in part to the high complexity of many solutions in economic models under this assumption. In response, models of bounded rationality and learning have recently flourished in economics, though the study and application of these ideas to approximate solutions of stochastic dynamic programming problems is still an emerging area. In particular, the study of consumption-saving decisions under uncertainty and liquidity constraints has been pursued by only a couple of papers with limited or negative results.

In this paper we have proposed an adaptive algorithm based on Euler equations and have studied its behavior through time using simulations. We have shown that this algorithm allows agents to have low welfare losses with high probability in a short time. In our simulations the probability of being in a neighborhood of the optimal rule is close to 1 within 500 periods for a set of commonly used parameterizations. Furthermore, we have generalized our adaptive procedure in order to allow for social learning and changes in the environment. In both cases the procedure had a similar behavior to our baseline model, with convergence of behavior to rules with low associated costs (less than 3%) with probabilities over 0.85 within 500 periods for most parametrizations. Additionally, we analyzed the effect of softer liquidity constraints on agent's behavior. As expected, we found that lower constraint levels induce agents to learn more slowly, since the cost of not behaving optimally falls. These results imply much faster learning than previous results in the literature, which were derived from learning algorithms based on estimation of the value function for different rules.

An additional feature of the adaptive theory we studied here is its continual out-of-equilibrium non-optimal dynamics at the individual level, where each agent's behavior never settles down and is non-optimal almost all the time. At the same time, aggregate behavior converges to an equilibrium, in which the average or median agent seems to be behaving (nearly) optimally. This property is missing in many macroeconomic models, in which both the system and the individual behave similarly, falling into a fallacy of composition. We have shown that very simple variations of the model can incorporate ideas of social learning and structural change, without generating an increased informational or computational burden for the agent. Furthermore, the theory is so simple that any sophisticated agent could use pen and paper to update their consumption rule in a way that will give her a higher level of welfare.

Even though our results seem promising, more research into this model and its extensions is required in order to see if such an approximation to the rational solution is learnable within the lifetime of one individual.[9] All the simulations were based on parameterizations, which implied an annual interpretation of time. It would be interesting to analyze the model under parameterizations that allowed for updating of the rule to occur on a monthly or quarterly basis. If the time for convergence is still of the same order, one could argue with confidence that this is a good model of adaptive consumption behavior.

If further exploration of the model proves successful, it would be a step in the direction of an alternative bottom-up approach to macroeconomics, along the lines advocated by Weintraub (1979), Leijonhufvud (1993) and others.[10] This bottom-up approach would endow agents not with decision rules that are always perfectly tailored to their specific environment, but rather with simple all-purpose rules that allow the agent to adapt in a plausibly opportunistic yet imperfect fashion to any given environment. This approach would allow us to ask how an economic system works to coordinate, for better or worse, the independent decisions of heterogeneous interacting agents; a question that the more conventional top-down approach evades by restricting attention to equilibrium states. We need a bottom-up approach in order to understand how the "invisible hand" works in actual free market systems, to diagnose what has gone wrong when coordination failure leads to macroeconomic crisis, and to prescribe system changes that reduce the likelihood of such failures.

[9] Özak (2009) presents analytical results on the behavior of the model for both linear and polynomial consumption rules.

[10] In particular, Weintraub (1979) argues that "a successful reconciliation of micro and macro might entail a return to Marshallian price theory, or a well worked out statement of individual behavior in a non-optimizing framework" (Weintraub, 1979, p.157).

Appendices

A Tables and Figures

Table 1: EV_*^*, EC_*^*, EV_γ^* and EC_γ^* under optimal linear rule, for each configuration of δ and θ

δ	θ	EV_*^*	EC_*^*	EV_γ^*	EC_γ^*
0.9	1.5	-0.1607	0.9841	-0.1650	0.9837
	2	-0.1731	0.9830	-0.1786	0.9824
	3	-0.2001	0.9806	-0.2027	0.9803
	3.5	-0.2128	0.9795	-0.2220	0.9786
	4	-0.2257	0.9784	-0.2275	0.9782
0.95	1.5	-0.1962	0.9903	-0.1987	0.9901
	2	-0.2184	0.9892	-0.2213	0.9891
	3	-0.2555	0.9875	-0.2554	0.9875
	3.5	-0.2709	0.9868	-0.2819	0.9862
	4	-0.2856	0.9861	-0.2848	0.9862

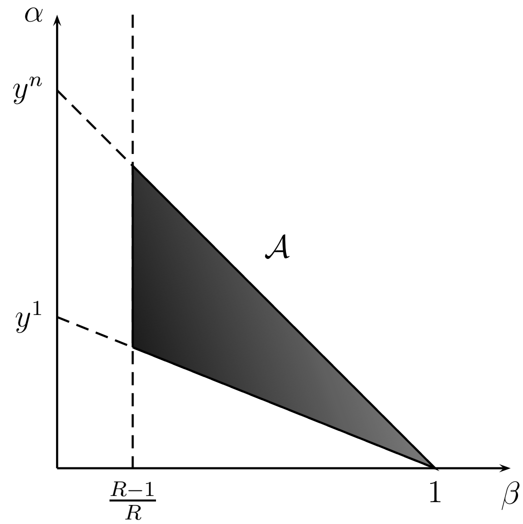

Figure 1: Admissible set of parameters.

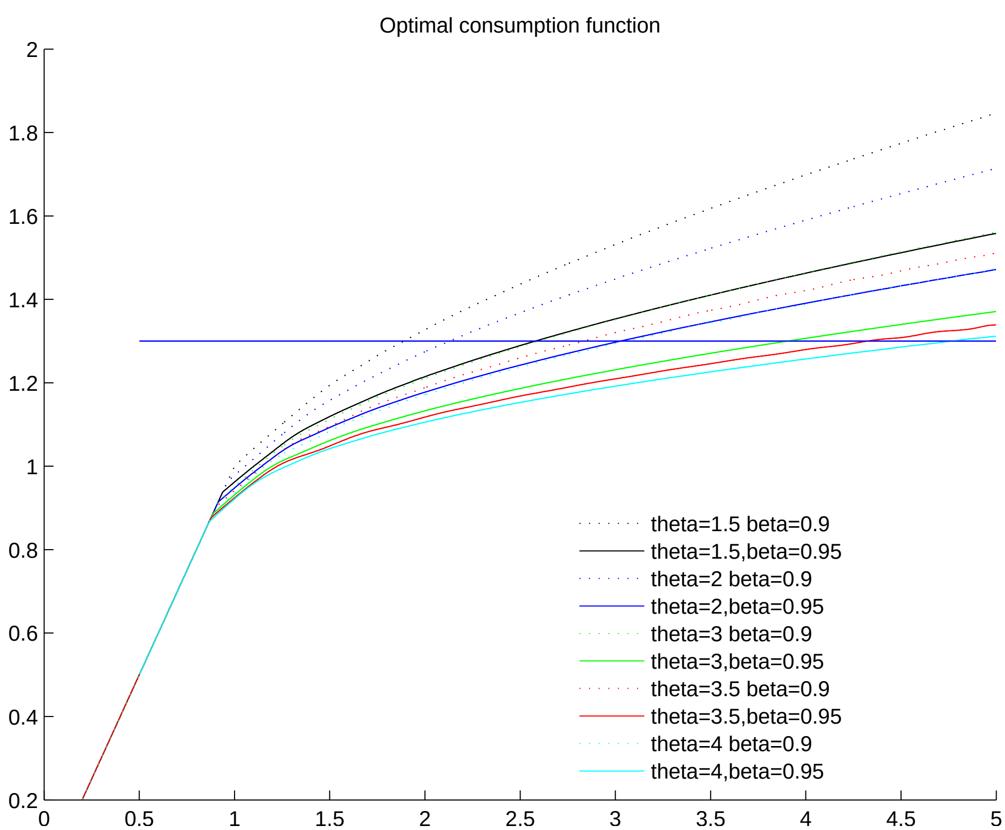

Figure 2: Optimal Consumption function for different parametrizations

Table 2: EV_*^γ, EC_*^γ, EV_γ^γ and EC_γ^γ under optimal linear rule, for each configuration of δ and θ

δ	θ	EV_*^γ	EC_*^γ	EV_γ^γ	EC_γ^γ
0.9	1.5	-0.1961	0.9807	-0.2003	0.9803
	2	-0.2092	0.9795	-0.2148	0.9790
	3	-0.2382	0.9770	-0.2408	0.9768
	3.5	-0.2515	0.9759	-0.2607	0.9751
	4	-0.2662	0.9747	-0.2680	0.9746
0.95	1.5	-0.2705	0.9866	-0.2730	0.9865
	2	-0.2943	0.9855	-0.2972	0.9854
	3	-0.3349	0.9837	-0.3347	0.9837
	3.5	-0.3525	0.9829	-0.3635	0.9824
	4	-0.3319	0.9839	-0.3310	0.9840

Table 3: Percentage difference between between equivalent consumption measures, $D_1^\gamma = \frac{EC_*^* - EC_*^\gamma}{EC_*^*} * 100$ and $D_2^\gamma = \frac{EC_\gamma^* - EC_\gamma^\gamma}{EC_\gamma^*} * 100$, and optimal linear rule. Percentage loss under the consume everything rule and expected loss over the whole set of parameters and conditional on rule being in the admissible set.

		Optimal Rule				Consume Everything		Random Rule			
δ	θ	D_1^γ	D_2^γ	α^*	β^*	$D_1^{(0,1)}$	$D_2^{(0,1)}$	$E(D_1^\gamma)$	$E(D_1^\gamma \mid \mathcal{A})$	$E(D_2^\gamma)$	$E(D_2^\gamma \mid \mathcal{A})$
0.9	1.5	0.35	0.35	0.61	0.39	0.98	0.93	6.54	3.79	6.34	3.27
	2	0.35	0.35	0.62	0.36	1.54	1.38	7.00	4.19	6.61	3.48
	3	0.36	0.36	0.61	0.34	3.00	2.50	8.16	5.28	7.38	4.16
	3.5	0.37	0.36	0.65	0.30	3.84	3.11	8.80	5.90	7.82	4.55
	4	0.38	0.38	0.65	0.29	4.71	3.74	9.49	6.57	8.28	4.98
0.95	1.5	0.37	0.37	0.63	0.33	1.57	1.45	5.78	3.09	5.64	2.77
	2	0.37	0.37	0.66	0.29	2.34	2.10	6.41	3.65	6.14	3.20
	3	0.38	0.38	0.71	0.23	4.09	3.52	7.84	4.96	7.28	4.21
	3.5	0.39	0.39	0.73	0.21	5.03	4.26	8.60	5.68	7.89	4.75
	4	0.22	0.22	0.72	0.21	5.79	4.83	6.97	5.57	4.94	4.19

Table 4: Probability of $D_1^{\gamma t}$ or $D_2^{\gamma t}$ below 0.5% at different periods for different parametrizations of the algorithm for $\eta = 1$, $\xi = 1$ and $\varepsilon = 0$.

		$t = 0$		$t = 50$		$t = 100$		$t = 250$		$t = 500$	
δ	θ	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$
0.9	1.5	0.2863	0.2950	0.8980	0.8943	0.9650	0.9581	0.9954	0.9927	0.9997	0.9993
	2	0.1410	0.1444	0.7706	0.7748	0.9181	0.9149	0.9847	0.9791	0.9963	0.9945
	3	0.0127	0.0127	0.4573	0.4390	0.6538	0.6232	0.8781	0.8398	0.9644	0.9424
	3.5	0.0104	0.0107	0.4214	0.4204	0.6241	0.6188	0.8678	0.8577	0.9544	0.9490
	4	0.0082	0.0084	0.3706	0.3695	0.5725	0.5667	0.8300	0.8184	0.9405	0.9290
0.95	1.5	0.0659	0.0667	0.7436	0.7461	0.9017	0.9021	0.9772	0.9772	0.9877	0.9877
	2	0.0155	0.0155	0.5636	0.5614	0.7575	0.7516	0.9212	0.9159	0.9675	0.9649
	3	0.0080	0.0086	0.3535	0.3643	0.5453	0.5575	0.7741	0.7827	0.8909	0.8972
	3.5	0.0075	0.0079	0.2739	0.2841	0.4241	0.4376	0.6544	0.6706	0.8003	0.8138
	4	0.0159	0.0176	0.4773	0.5009	0.6622	0.6913	0.8552	0.8736	0.9248	0.9319

Table 5: Probability of $D_1^{\gamma t}$ or $D_2^{\gamma t}$ below 0.5% at different periods for different parametrizations of the algorithm for $\theta = 3.5$ and $\delta = 0.95$.

			$t = 0$		$t = 50$		$t = 100$		$t = 250$		$t = 500$	
η	ξ	ε	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$	$D_1^{\gamma t}$	$D_2^{\gamma t}$
0.5	0	0.2	0.0075	0.0079	0.2730	0.2783	0.2497	0.2520	0.2284	0.2307	0.2223	0.2243
0.5	0	0	0.0075	0.0079	0.2222	0.2329	0.2919	0.3057	0.3999	0.4193	0.4970	0.5167
0.5	1	0.2	0.0075	0.0079	0.2549	0.2627	0.2893	0.2981	0.2922	0.3000	0.2933	0.3011
0.5	1	0	0.0075	0.0079	0.1468	0.1519	0.1954	0.2019	0.2799	0.2888	0.3660	0.3750
1	0	0.2	0.0075	0.0079	0.1355	0.1395	0.1267	0.1303	0.1203	0.1230	0.1216	0.1243
1	0	0	0.0075	0.0079	0.2234	0.2328	0.3355	0.3497	0.5153	0.5344	0.6404	0.6561
1	1	0.2	0.0075	0.0079	0.1273	0.1324	0.1262	0.1309	0.1265	0.1316	0.1279	0.1324
1	1	0	0.0075	0.0079	0.2739	0.2841	0.4241	0.4376	0.6544	0.6706	0.8003	0.8138

Table 6: Probability of $D_1^{\gamma_t}$ or $D_2^{\gamma_t}$ below 0.5% at different periods for different parametrizations of the algorithm under social learning for $\phi = \frac{1}{3}$, $\theta = 3.5$ and $\delta = 0.95$.

Indiv=200, Neigh=2			$t = 0$		$t = 50$		$t = 100$		$t = 250$		$t = 500$	
η	ξ	ε	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$
0.5	0	0.2	0.0075	0.0079	0.5216	0.5216	0.2787	0.2787	0.1790	0.1790	0.1841	0.1841
0.5	0	0	0.0075	0.0079	0.6502	0.6620	0.8203	0.8243	0.9606	0.9606	0.9939	0.9939
0.5	1	0.2	0.0075	0.0079	0.5980	0.5997	0.6468	0.6475	0.6531	0.6538	0.6380	0.6389
0.5	1	0	0.0075	0.0079	0.0276	0.0278	0.0549	0.0549	0.1800	0.1800	0.2898	0.2898
1	0	0.2	0.0075	0.0079	0.4351	0.4356	0.3520	0.3522	0.3308	0.3308	0.3397	0.3397
1	0	0	0.0075	0.0079	0.8995	0.9020	0.9848	0.9849	0.9992	0.9992	1.0000	1.0000
1	1	0.2	0.0075	0.0079	0.4810	0.4875	0.4794	0.4862	0.4850	0.4910	0.4760	0.4826
1	1	0	0.0075	0.0079	0.7925	0.7960	0.9513	0.9513	0.9991	0.9991	1.0000	1.0000

Table 7: Expected loss under optimal distribution, $ED_1^{\gamma_t}$, and under linear rule's distribution, $ED_2^{\gamma_t}$, at different periods for different parametrizations of the consumer's problem for $\eta = 1$, $\xi = 1$ and $\varepsilon = 0$.

		$t = 0$		$t = 5$		$t = 10$		$t = 25$		$t = 50$	
δ	θ	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$
0.9	1.5	1.20	1.38	1.09	1.26	0.77	0.85	0.48	0.50	0.42	0.42
	2	1.33	1.43	1.35	1.48	0.96	1.04	0.58	0.59	0.47	0.48
	3	2.12	2.04	1.97	1.97	1.42	1.44	0.81	0.81	0.60	0.60
	3.5	2.65	2.46	2.33	2.26	1.70	1.68	0.94	0.94	0.67	0.66
	4	3.24	2.93	2.72	2.58	2.00	1.94	1.10	1.08	0.75	0.74
0.95	1.5	1.26	1.28	1.25	1.30	0.90	0.93	0.58	0.58	0.49	0.49
	2	1.72	1.69	1.59	1.60	1.12	1.13	0.69	0.69	0.56	0.55
	3	2.91	2.70	2.33	2.25	1.69	1.65	0.95	0.93	0.71	0.70
	3.5	3.61	3.29	2.79	2.65	2.01	1.94	1.12	1.08	0.80	0.78
	4	4.06	3.57	2.96	2.63	2.13	1.94	1.11	1.04	0.73	0.69

Table 8: Median loss under optimal distribution, $MD_1^{\gamma_t}$, and linear rule's distribution, $MD_2^{\gamma_t}$, at different periods for different parametrizations of the consumer's problem for $\eta = 1$, $\xi = 1$ and $\varepsilon = 0$.

		$t=0$		$t=5$		$t=10$		$t=25$		$t=50$	
δ	θ	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$
0.9	1.5	0.67	0.65	0.67	0.65	0.55	0.54	0.41	0.42	0.39	0.38
	2	0.98	0.92	0.96	0.91	0.73	0.71	0.49	0.48	0.43	0.43
	3	2.02	1.79	1.68	1.54	1.14	1.10	0.65	0.65	0.51	0.52
	3.5	2.68	2.32	2.10	1.89	1.38	1.30	0.73	0.74	0.54	0.54
	4	3.36	2.86	2.54	2.26	1.65	1.55	0.82	0.81	0.58	0.58
0.95	1.5	1.09	1.04	0.94	0.90	0.69	0.69	0.49	0.49	0.44	0.44
	2	1.66	1.55	1.27	1.22	0.89	0.87	0.57	0.57	0.48	0.48
	3	3.04	2.74	2.06	1.92	1.33	1.28	0.75	0.73	0.57	0.57
	3.5	3.84	3.42	2.52	2.32	1.59	1.50	0.85	0.83	0.63	0.62
	4	4.42	3.88	2.79	2.52	1.69	1.57	0.78	0.74	0.52	0.50

Table 9: Expected loss under optimal distribution, $ED_1^{\gamma_t}$, and under linear rule's distribution, $ED_2^{\gamma_t}$, at different periods for different parametrizations of the algorithm for $\theta = 3.5$ and $\delta = 0.95$.

			$t = 0$		$t = 5$		$t = 10$		$t = 25$		$t = 50$	
η	ξ	ε	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$
0.5	0	0.2	3.61	3.29	2.51	2.30	1.74	1.63	1.05	1.01	0.94	0.93
0.5	0	0	3.61	3.29	2.49	2.30	1.71	1.60	1.12	1.07	0.92	0.89
0.5	1	0.2	3.61	3.29	2.73	2.47	2.02	1.85	1.14	1.09	0.81	0.80
0.5	1	0	3.61	3.29	2.75	2.49	2.13	1.95	1.47	1.38	1.12	1.07
1	0	0.2	3.61	3.29	2.79	2.82	2.67	2.86	2.01	2.07	1.95	2.05
1	0	0	3.61	3.29	2.79	2.81	2.38	2.51	1.51	1.49	1.28	1.26
1	1	0.2	3.61	3.29	2.79	2.65	2.17	2.13	1.40	1.42	1.24	1.25
1	1	0	3.61	3.29	2.79	2.65	2.01	1.94	1.12	1.08	0.80	0.78

Table 10: Median loss under optimal distribution, $MD_1^{\gamma_t}$, and under linear rule's distribution, $MD_2^{\gamma_t}$, at different periods for different parametrizations of the algorithm for $\theta = 3.5$ and $\delta = 0.95$.

			$t = 0$		$t = 5$		$t = 10$		$t = 25$		$t = 50$	
η	ξ	ε	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$	$MD_1^{\gamma_t}$	$MD_2^{\gamma_t}$
0.5	0	0.2	3.84	3.42	2.46	2.28	1.39	1.33	0.75	0.74	0.67	0.66
0.5	0	0	3.84	3.42	2.46	2.28	1.31	1.26	0.82	0.79	0.69	0.67
0.5	1	0.2	3.84	3.42	2.84	2.59	1.82	1.69	0.85	0.83	0.66	0.65
0.5	1	0	3.84	3.42	2.87	2.62	1.97	1.84	1.21	1.16	0.92	0.89
1	0	0.2	3.84	3.42	2.13	1.98	1.69	1.61	1.10	1.08	1.00	1.00
1	0	0	3.84	3.42	2.13	2.01	1.52	1.45	0.91	0.88	0.70	0.68
1	1	0.2	3.84	3.42	2.52	2.32	1.72	1.62	1.06	1.03	0.98	0.96
1	1	0	3.84	3.42	2.52	2.32	1.59	1.50	0.85	0.83	0.63	0.62

Table 11: Expected loss under optimal distribution, $ED_1^{\gamma_t}$, and under linear rule's distribution, $ED_2^{\gamma_t}$, at different periods for different parametrizations of the algorithm under social learning for $\phi = \frac{1}{3}$, $\theta = 3.5$ and $\delta = 0.95$.

Indiv=200, Neigh=2			$t=0$		$t=5$		$t=10$		$t=25$		$t=50$	
η	ξ	ε	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$	$ED_1^{\gamma_t}$	$ED_2^{\gamma_t}$
0.5	0	0.2	3.61	3.29	2.30	2.11	1.09	1.03	0.54	0.54	0.58	0.59
0.5	0	0	3.61	3.29	2.31	2.11	1.09	1.03	0.57	0.56	0.49	0.48
0.5	1	0.2	3.61	3.29	2.44	2.22	1.52	1.42	0.69	0.68	0.52	0.52
0.5	1	0	3.61	3.29	2.45	2.23	1.63	1.52	1.00	0.97	0.77	0.75
1	0	0.2	3.61	3.29	1.65	1.53	0.97	0.93	0.64	0.63	0.71	0.71
1	0	0	3.61	3.29	1.66	1.54	0.88	0.85	0.50	0.49	0.44	0.44
1	1	0.2	3.61	3.29	1.69	1.56	0.99	0.95	0.61	0.60	0.58	0.58
1	1	0	3.61	3.29	1.71	1.57	0.92	0.88	0.54	0.53	0.46	0.46

Table 12: Median loss under optimal distribution, $MD_1^{\gamma t}$, and under linear rule's distribution, $MD_2^{\gamma t}$, at different periods for different parametrizations of the algorithm under social learning for $\phi = \frac{1}{3}$, $\theta = 3.5$ and $\delta = 0.95$.

Indiv=200, Neigh=2			$t=0$		$t=5$		$t=10$		$t=25$		$t=50$	
η	ξ	ε	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$	$MD_1^{\gamma t}$	$MD_2^{\gamma t}$
0.5	0	0.2	3.84	3.42	2.39	2.21	0.91	0.87	0.46	0.46	0.48	0.48
0.5	0	0	3.84	3.42	2.39	2.22	0.92	0.87	0.51	0.51	0.47	0.47
0.5	1	0.2	3.84	3.42	2.54	2.32	1.46	1.37	0.63	0.62	0.47	0.47
0.5	1	0	3.84	3.42	2.55	2.34	1.59	1.49	0.96	0.93	0.75	0.73
1	0	0.2	3.84	3.42	1.52	1.41	0.74	0.71	0.50	0.49	0.52	0.53
1	0	0	3.84	3.42	1.52	1.42	0.66	0.65	0.46	0.45	0.42	0.42
1	1	0.2	3.84	3.42	1.57	1.44	0.81	0.78	0.53	0.53	0.51	0.51
1	1	0	3.84	3.42	1.58	1.45	0.75	0.73	0.50	0.49	0.45	0.45

Table 13: Percentage difference between between equivalent consumption measures, $D_1^\gamma = \frac{EC_*^* - EC_*^\gamma}{EC_*^*} * 100$ and $D_2^\gamma = \frac{EC_\gamma^* - EC_\gamma^\gamma}{EC_\gamma^*} * 100$, and optimal linear rule. Percentage loss under the consume everything rule and expected loss over the whole set of parameters and conditional on rule being in the admissible set for $\theta = 3.5$, $\delta = 0.95$, $B = 0$.

	Optimal Rule				Consume Everything		Random Rule			
Process	D_1^γ	D_2^γ	α^*	β^*	$D_1^{(0,1)}$	$D_2^{(0,1)}$	$E(D_1^\gamma)$	$E(D_1^\gamma \mid \mathcal{A})$	$E(D_2^\gamma)$	$E(D_2^\gamma \mid \mathcal{A})$
1	0.02	0.02	0.72	0.22	4.62	3.88	6.40	3.27	4.46	2.80
2	0.02	0.02	1.46	0.21	4.62	3.88	7.91	6.94	7.50	5.47
3	0.47	0.36	1.26	0.22	14.64	11.76	14.50	13.97	12.64	12.32
4	0.29	0.28	0.66	0.17	23.79	19.51	21.21	18.98	17.17	15.41
5	0.80	0.76	0.44	0.24	65.99	55.90	53.78	52.53	45.53	44.87

Table 14: Probability of $D_1^{\gamma_t}$ or $D_2^{\gamma_t}$ below 3% at different periods for fixed parametrization of the algorithm under different income processes for $\theta = 3.5$, $\delta = 0.95$, $B = 0$, $\eta = 1$, $\epsilon = 0$, $\xi = 1$ and $\phi = 1$.

	$t = 0$		$t = 50$		$t = 100$		$t = 250$		$t = 500$	
Process	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$
1	0.3846	0.4980	0.8952	0.8866	0.9064	0.9011	0.9194	0.9141	0.9291	0.9224
2	0.3860	0.4989	0.9054	0.8987	0.9156	0.9124	0.9290	0.9254	0.9373	0.9318
3	0.0252	0.0299	0.3394	0.3441	0.4497	0.4460	0.5427	0.5289	0.5885	0.5701
4	0.0120	0.0129	0.3537	0.3420	0.5006	0.4865	0.6472	0.6335	0.7040	0.6961
5	0.0031	0.0038	0.0462	0.0585	0.0785	0.0958	0.1423	0.1651	0.2142	0.2415

Table 15: Probability of $D_1^{\gamma_t}$ or $D_2^{\gamma_t}$ below 3% at different periods for fixed parametrization of the algorithm under different income processes under ID version of the algorithm for $\theta = 3.5$, $\delta = 0.95$, $B = 0$, $\eta = 1$, $\epsilon = 0$, $\xi = 1$ and $\phi = 1$.

	$t = 0$		$t = 50$		$t = 100$		$t = 250$		$t = 500$	
Process	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$	$D_1^{\gamma_t}$	$D_2^{\gamma_t}$
1	0.3846	0.4980	0.9680	0.9742	0.9735	0.9785	0.9804	0.9833	0.9847	0.9862
2	0.3860	0.4989	0.9678	0.9742	0.9728	0.9783	0.9790	0.9821	0.9832	0.9847
3	0.0252	0.0299	0.8103	0.8150	0.8495	0.8468	0.8707	0.8650	0.8840	0.8762
4	0.0120	0.0129	0.6666	0.7393	0.7615	0.8062	0.8317	0.8538	0.8590	0.8707
5	0.0031	0.0038	0.2321	0.2829	0.2652	0.3251	0.3024	0.3659	0.3109	0.3773

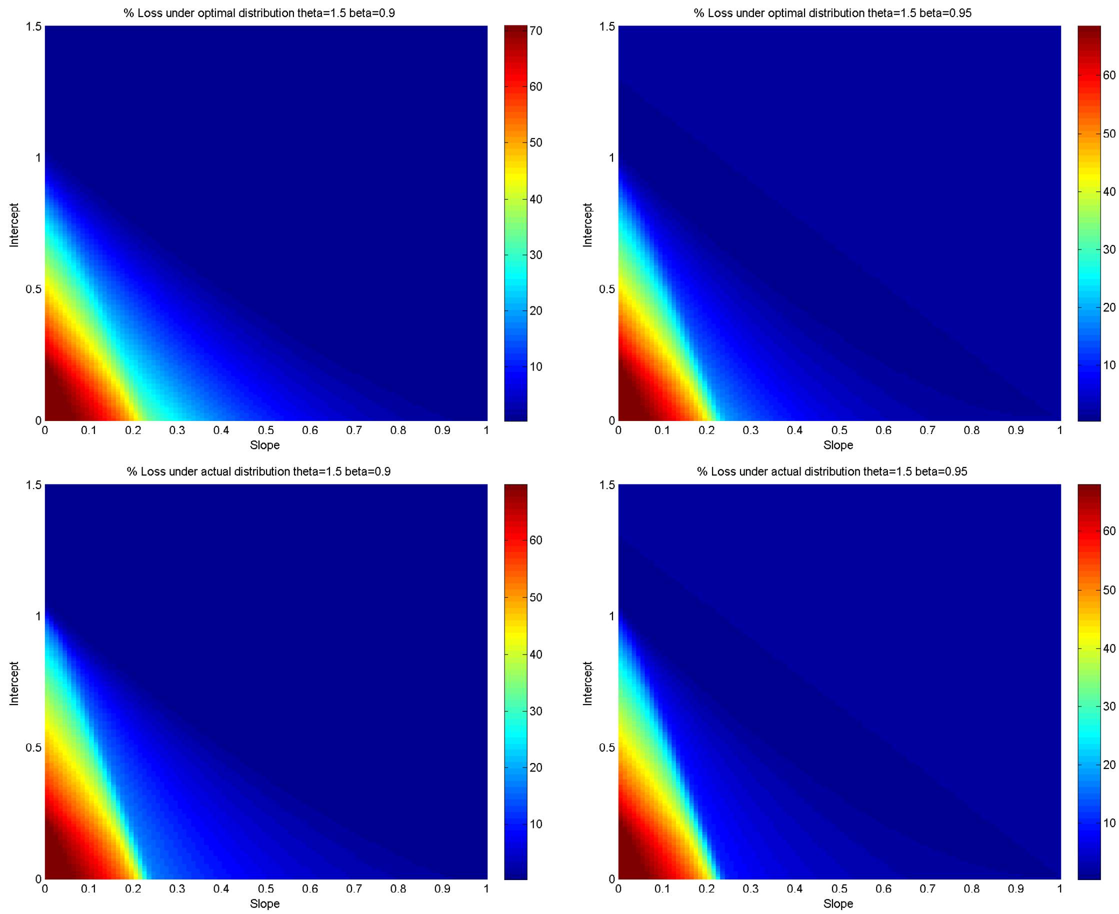

Figure 3: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept and slope, under π^* (above) and π^γ (below) for $\theta = 1.5$, $\delta = 0.9$ (left) and $\delta = 0.95$ (right).

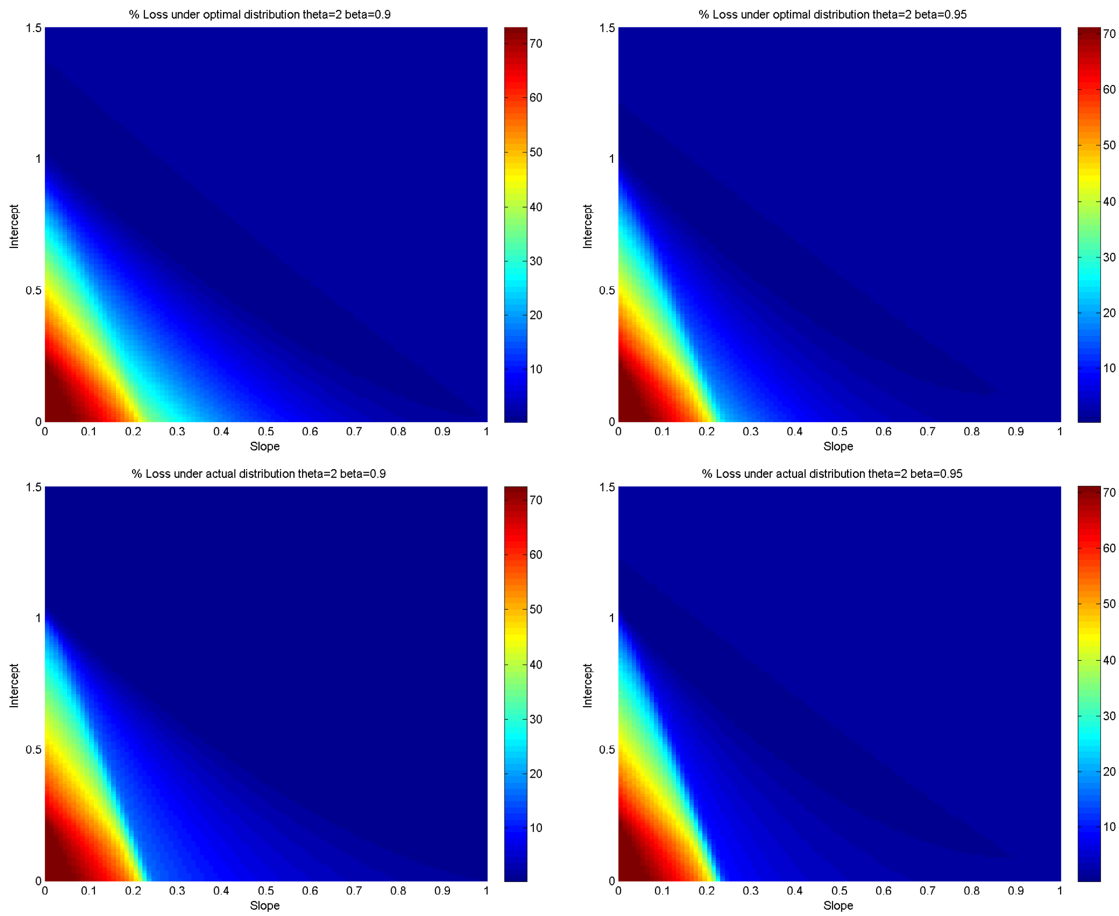

Figure 4: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept and slope, under π^* (above) and π^γ (below) for $\theta = 2$, $\delta = 0.9$ (left) and $\delta = 0.95$ (right).

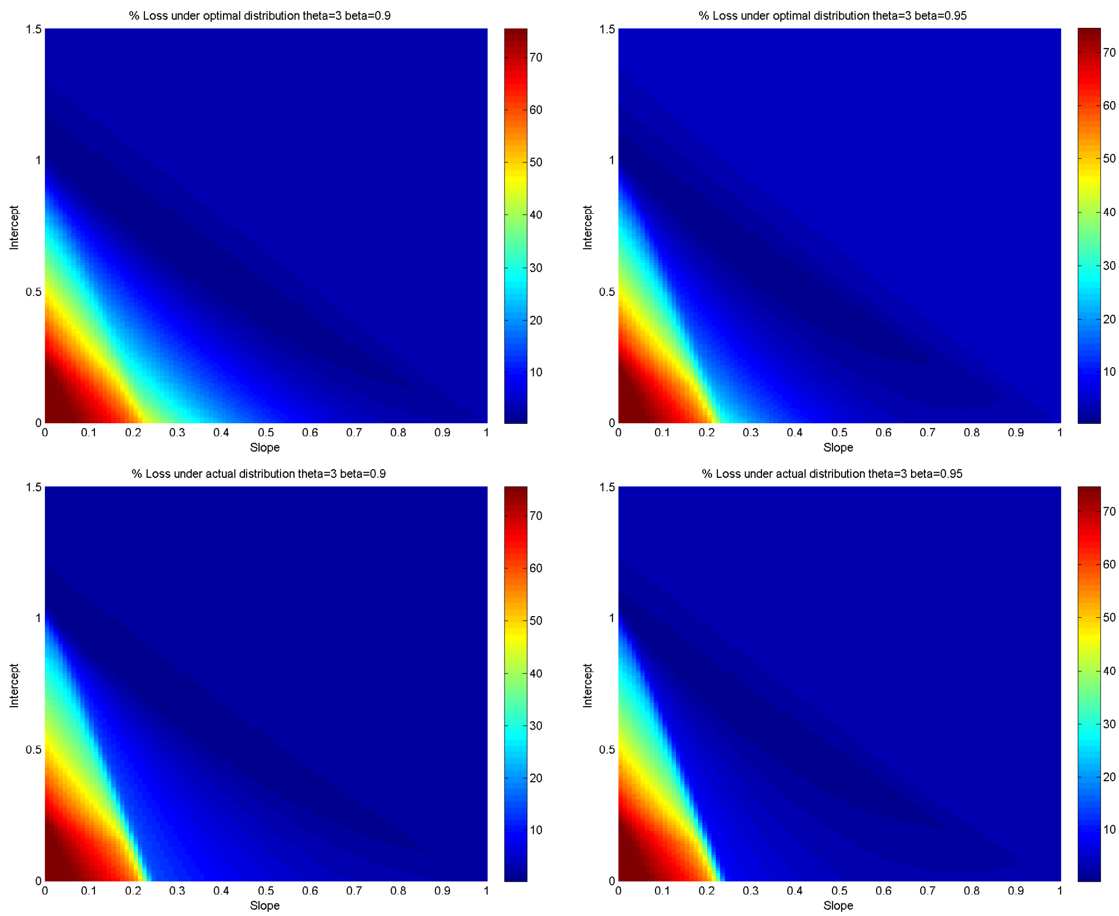

Figure 5: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept and slope, under π^* (above) and π^γ (below) for $\theta = 3$, $\delta = 0.9$ (left) and $\delta = 0.95$ (right).

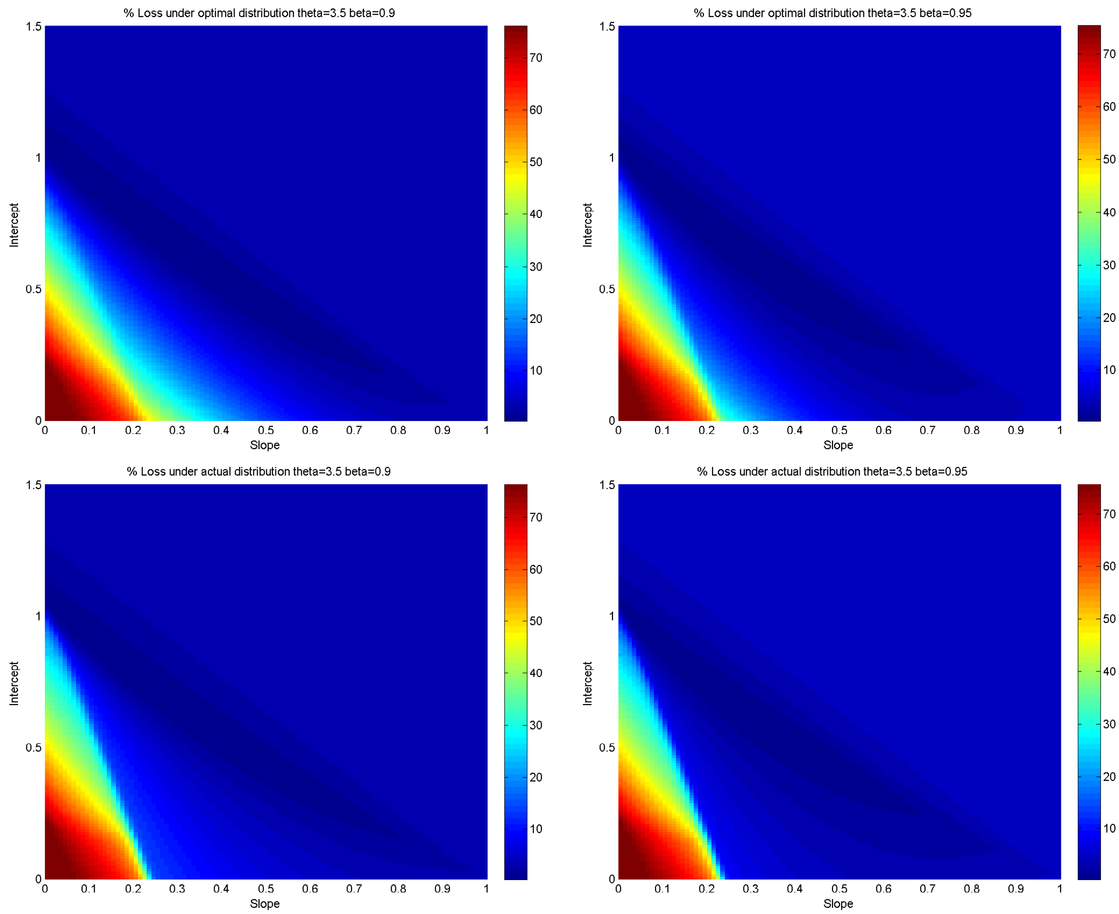

Figure 6: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept and slope, under π^* (above) and π^γ (below) for $\theta = 3.5$, $\delta = 0.9$ (left) and $\delta = 0.95$ (right).

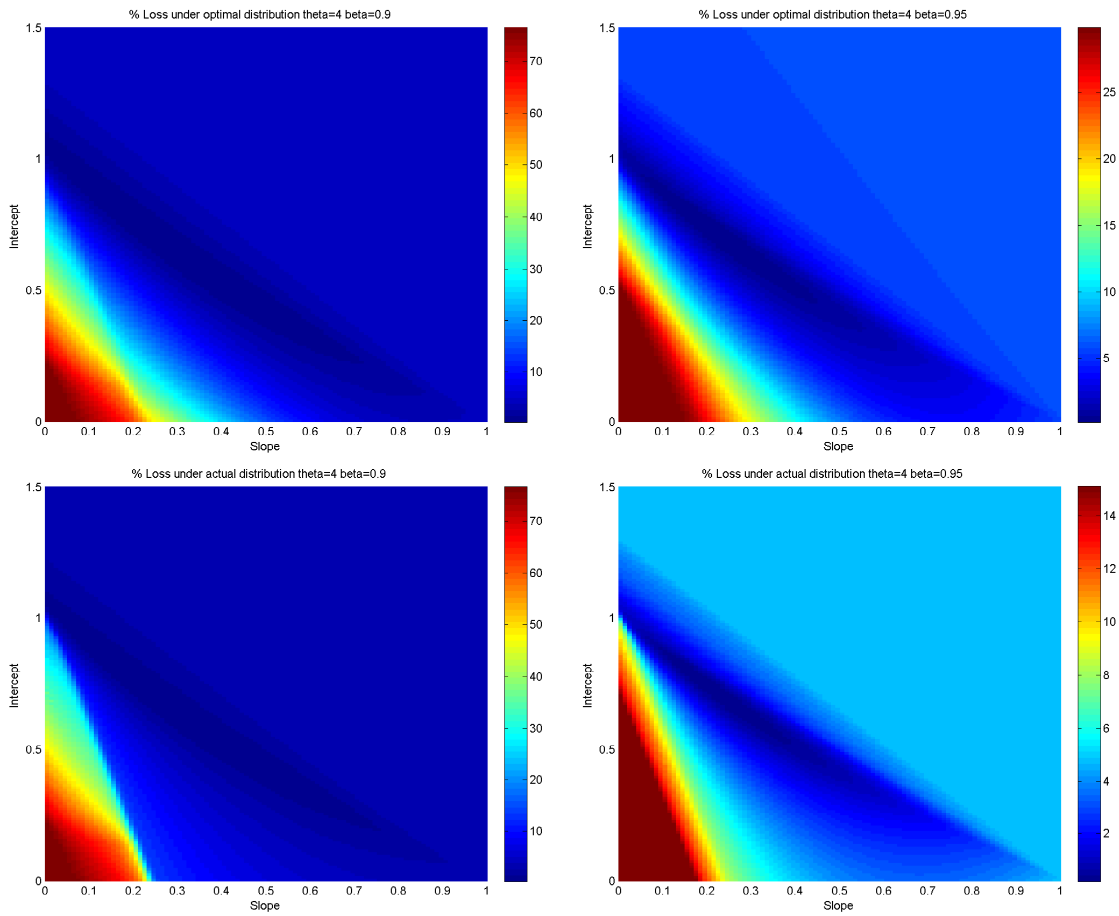

Figure 7: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept and slope, under π^* (above) and π^γ (below) for $\theta = 4$, $\delta = 0.9$ (left) and $\delta = 0.95$ (right).

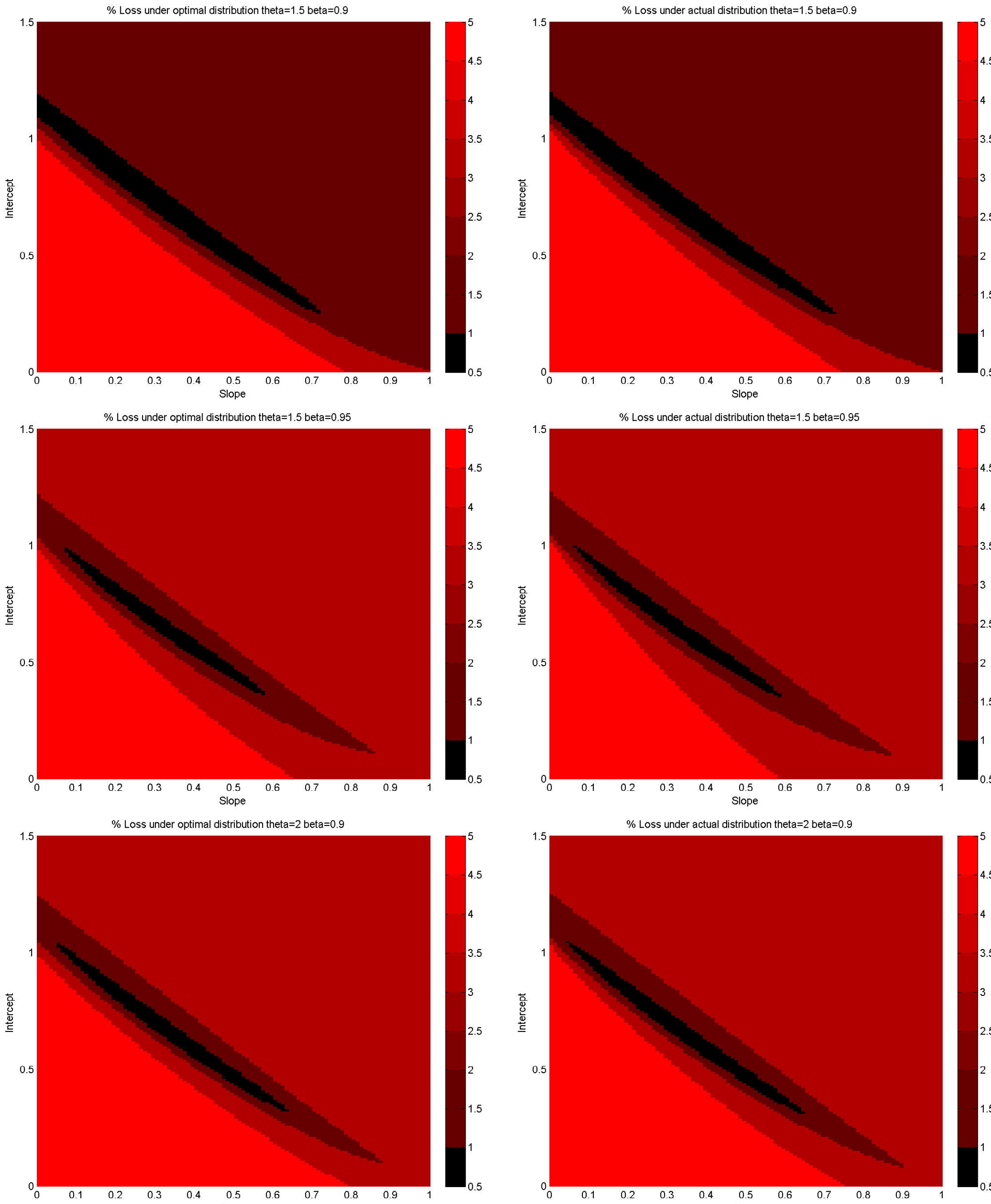

Figure 8: Set of consumption rules which have a percentage deviation less than or equal to 0.5%, 1%, 3% and more than 3%.

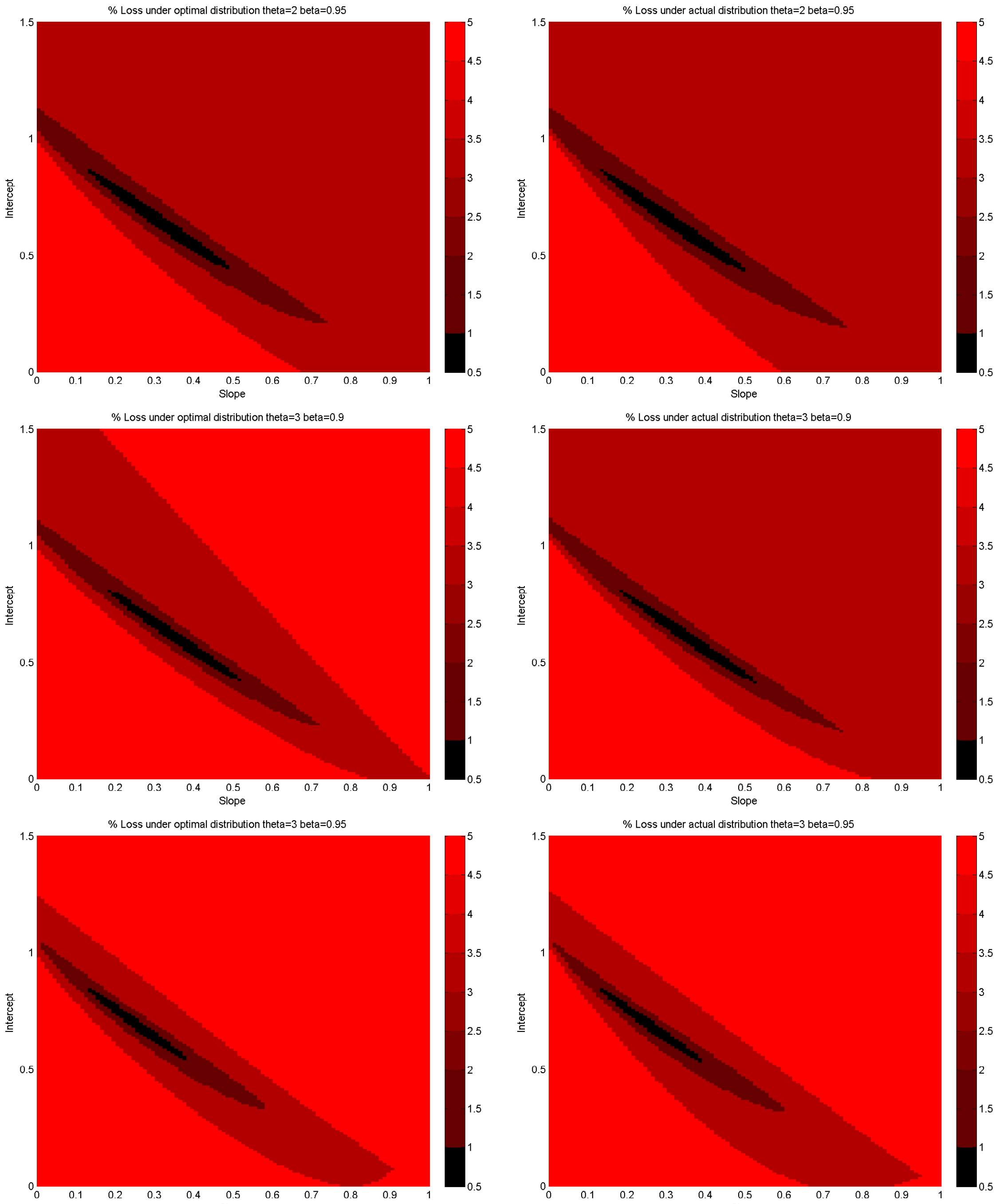

Figure 9: Set of consumption rules which have a percentage deviation less than or equal to 0.5% 1%, 3% and more than 3%.

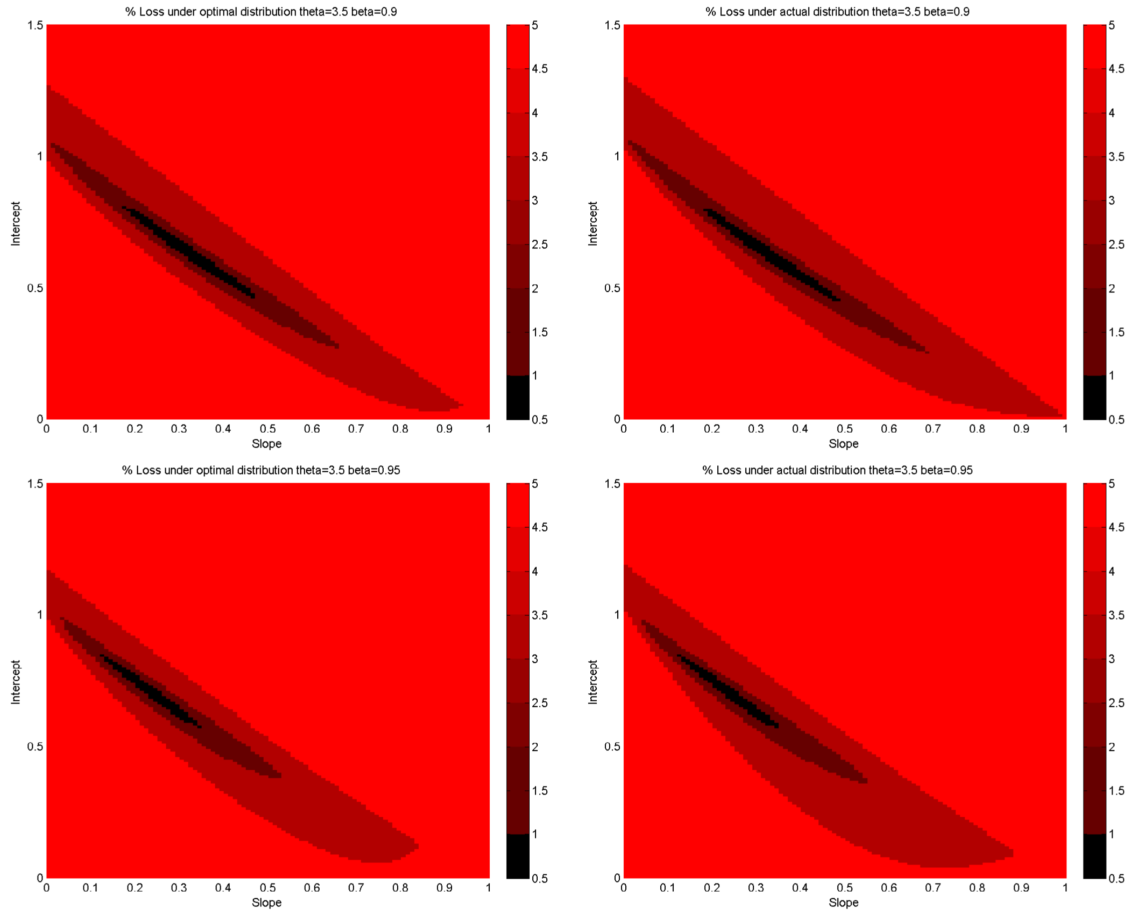

Figure 10: Set of consumption rules which have a percentage deviation less than or equal to 0.5% 1%, 3% and more than 3%.

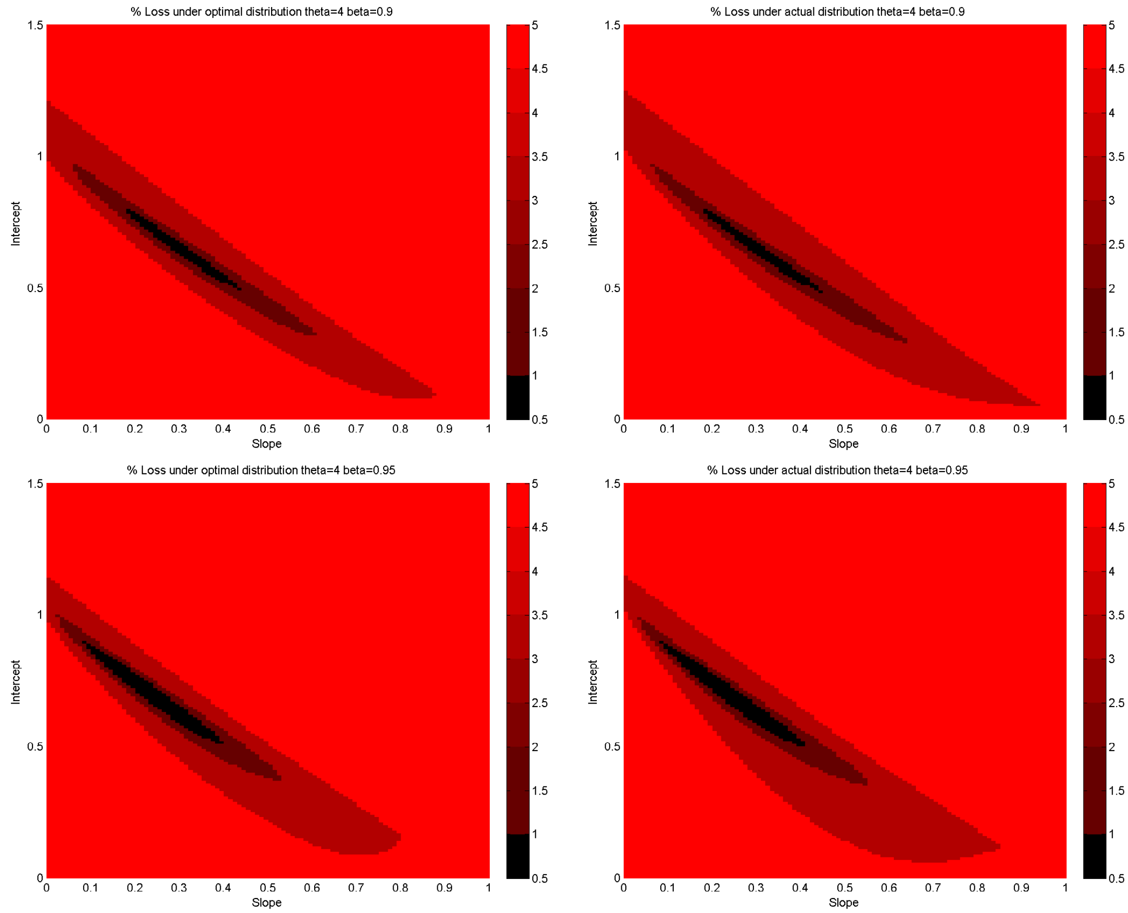

Figure 11: Set of consumption rules which have a percentage deviation less than or equal to 0.5% 1%, 3% and more than 3%.

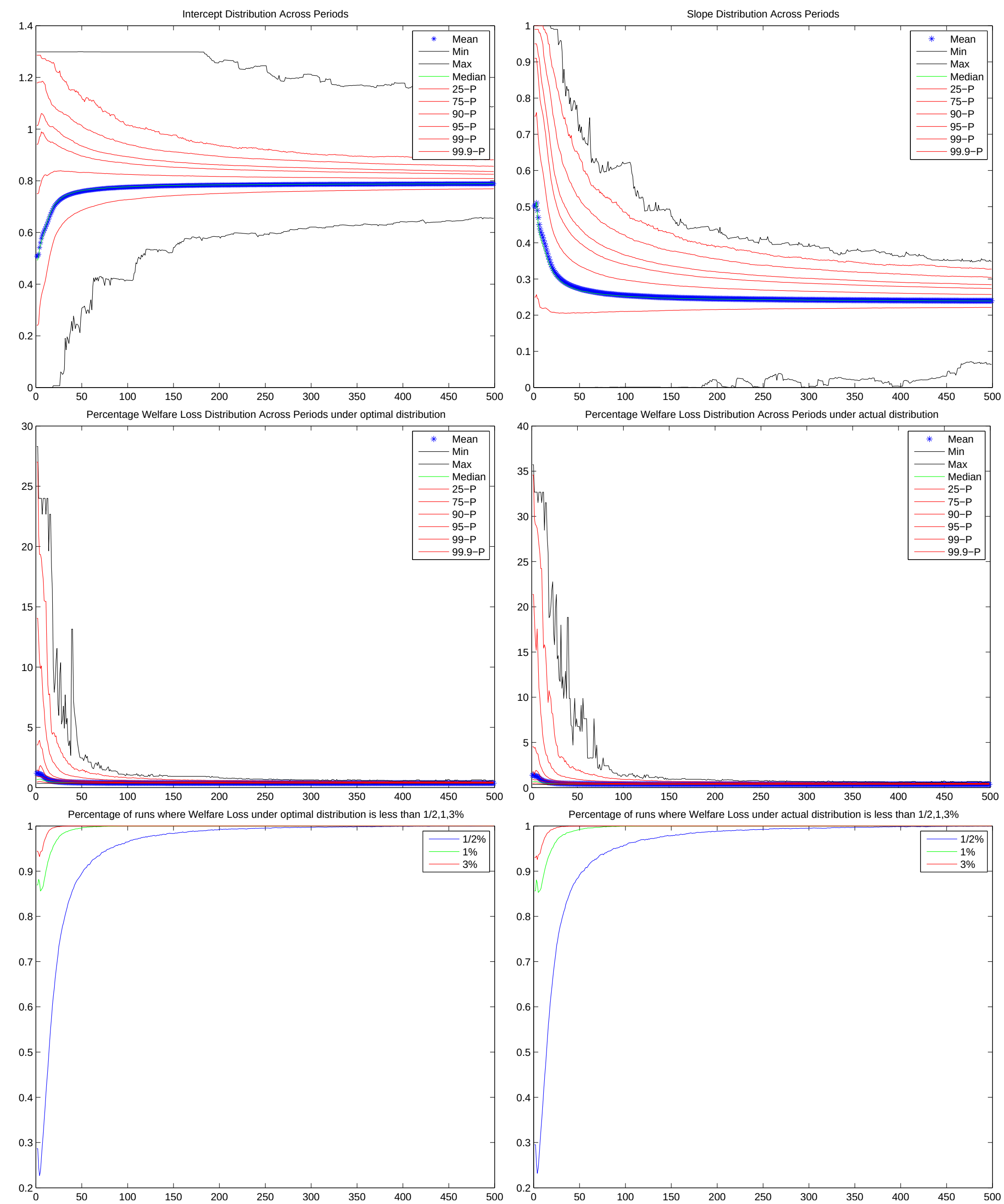

Figure 12: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 1.5$, $\delta = 0.9$.

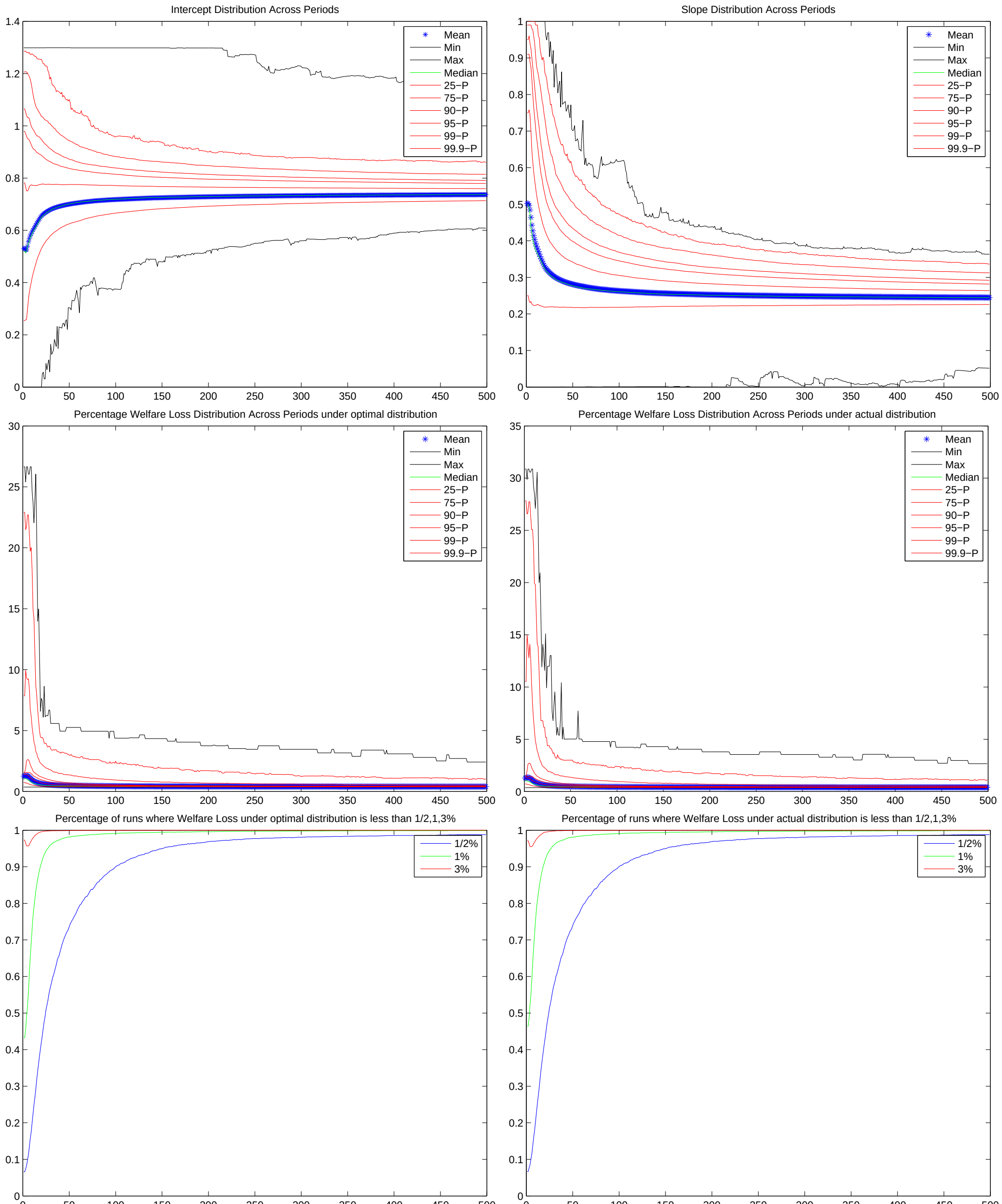

Figure 13: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 1.5$, $\delta = 0.95$.

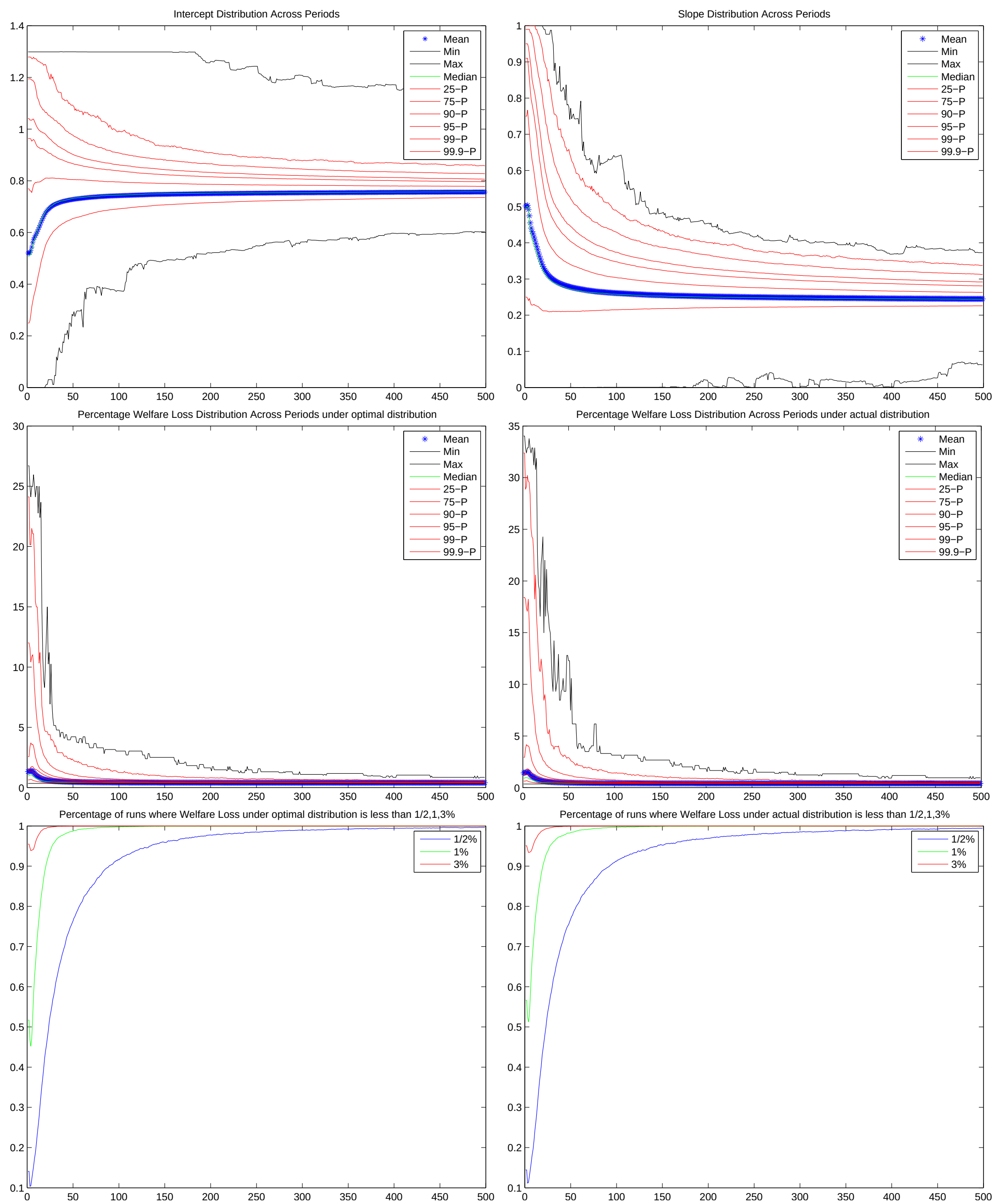

Figure 14: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 2$, $\delta = 0.9$.

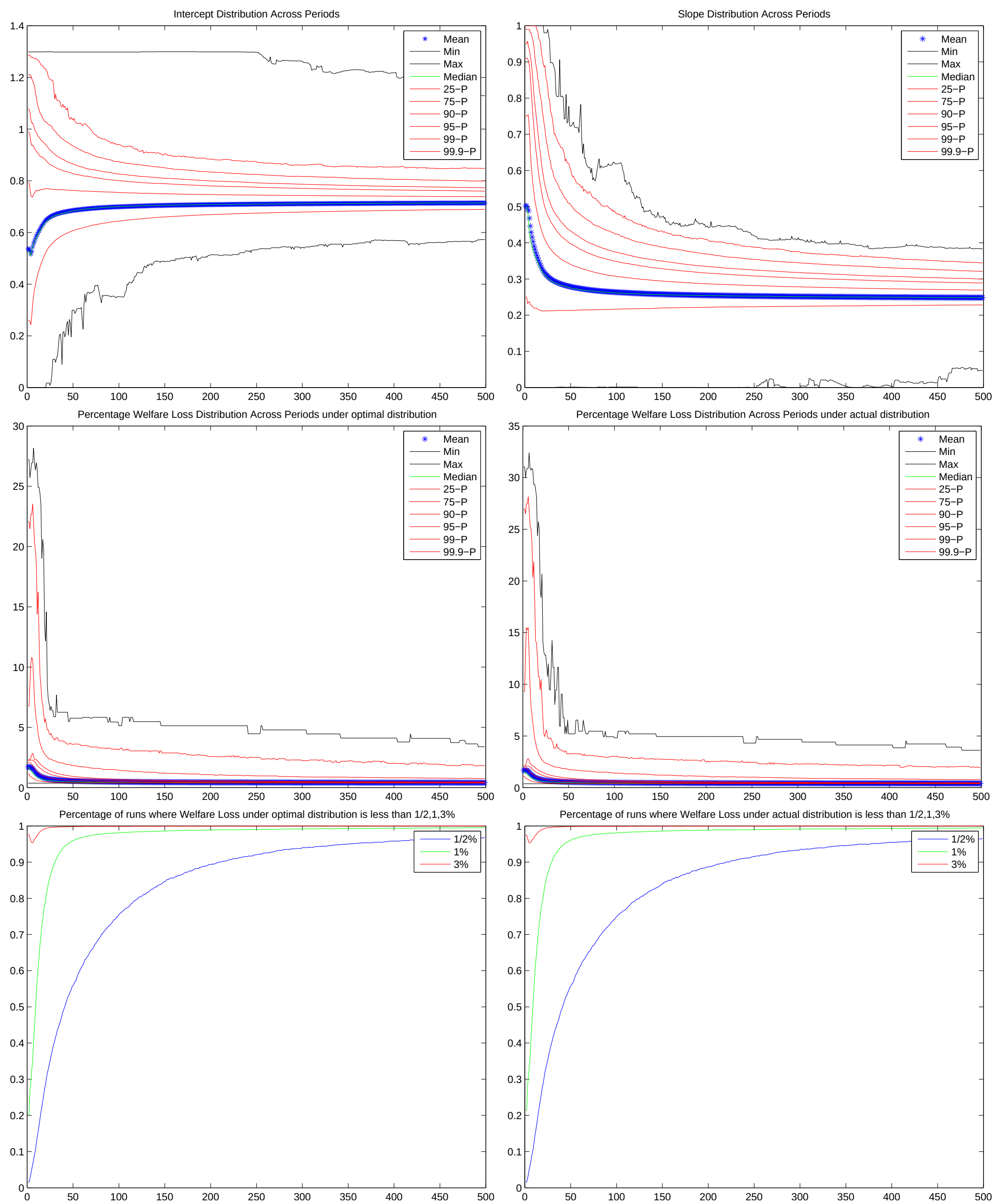

Figure 15: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 2$, $\delta = 0.95$.

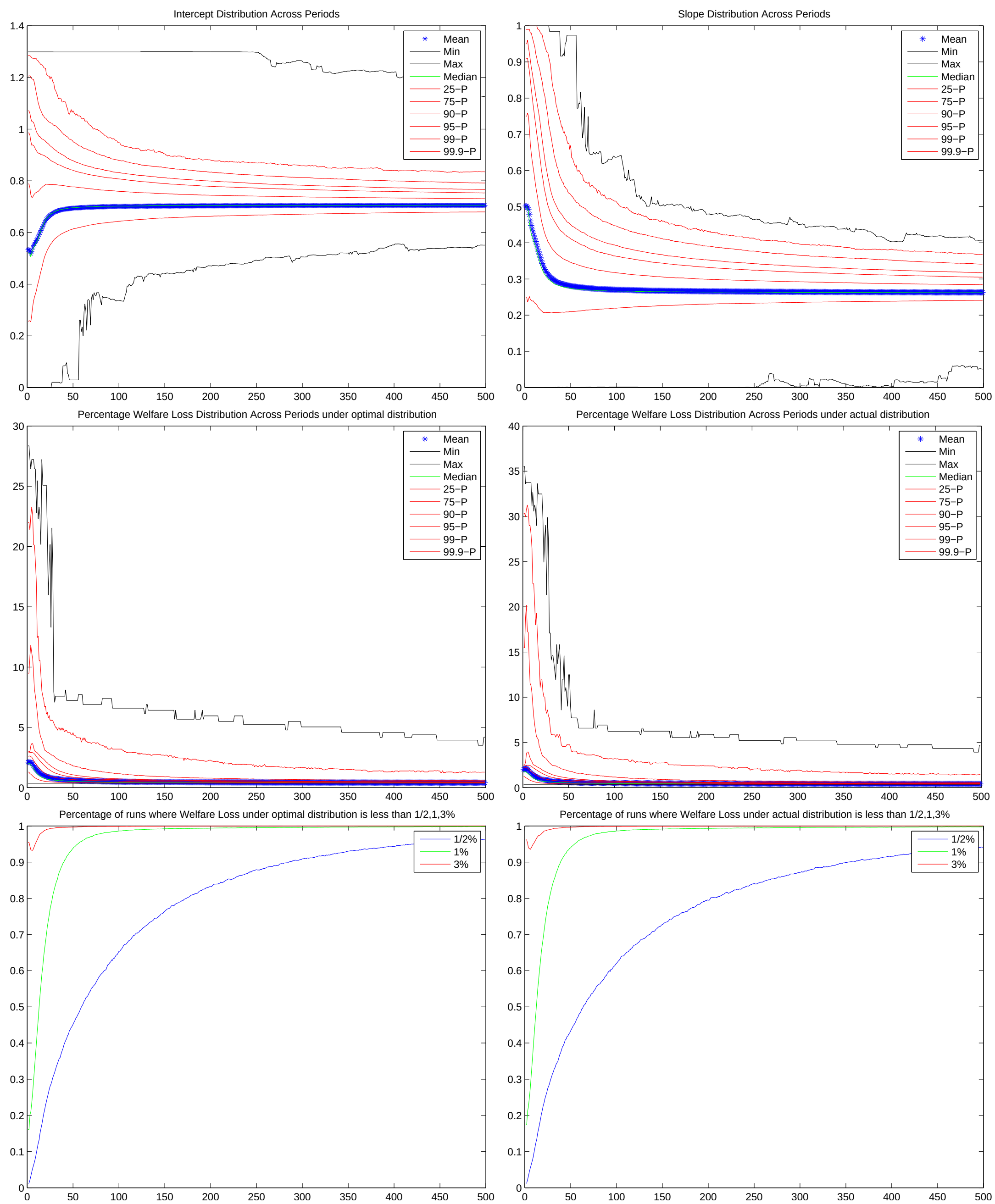

Figure 16: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 3$, $\delta = 0.9$.

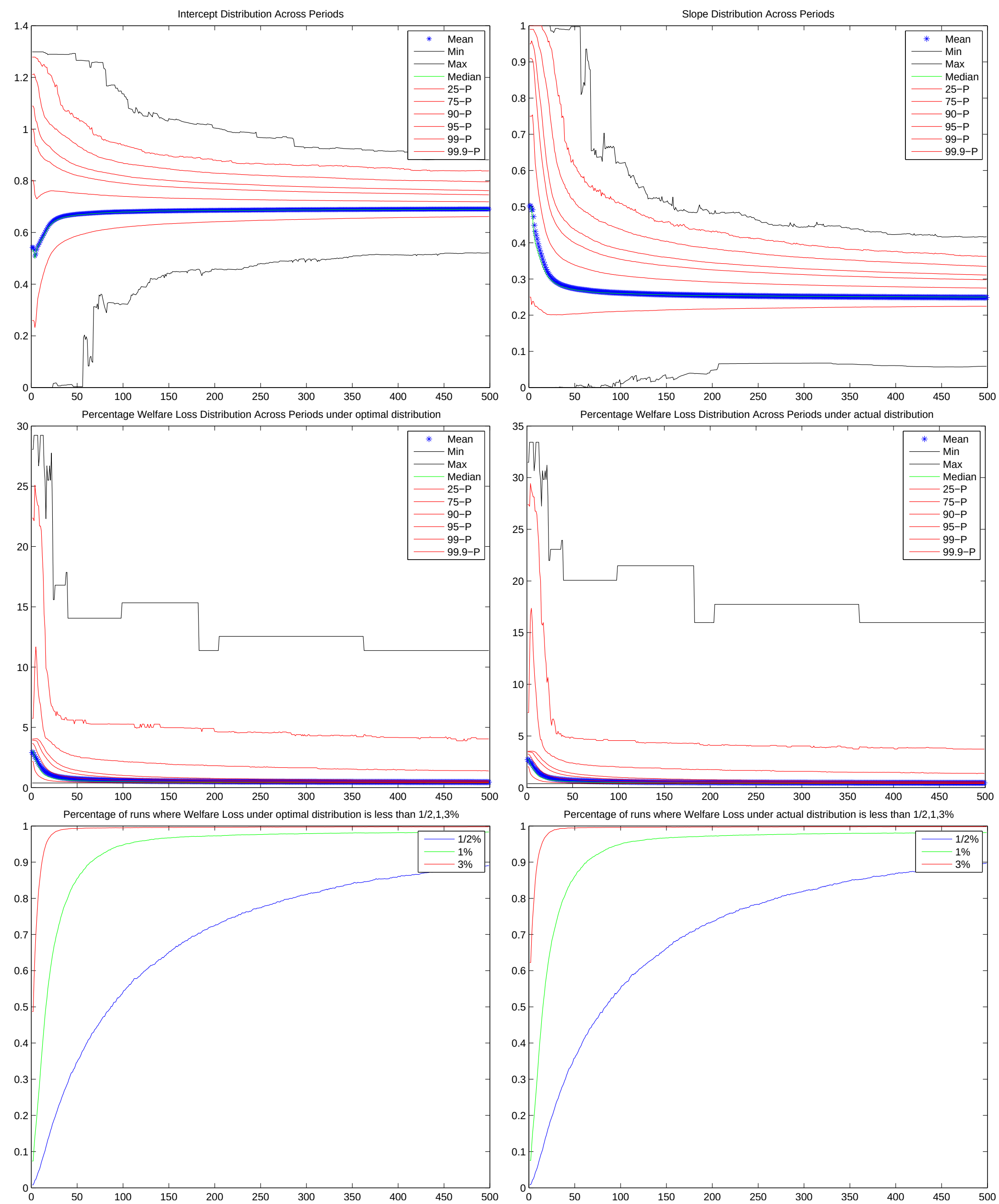

Figure 17: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 3$, $\delta = 0.95$.

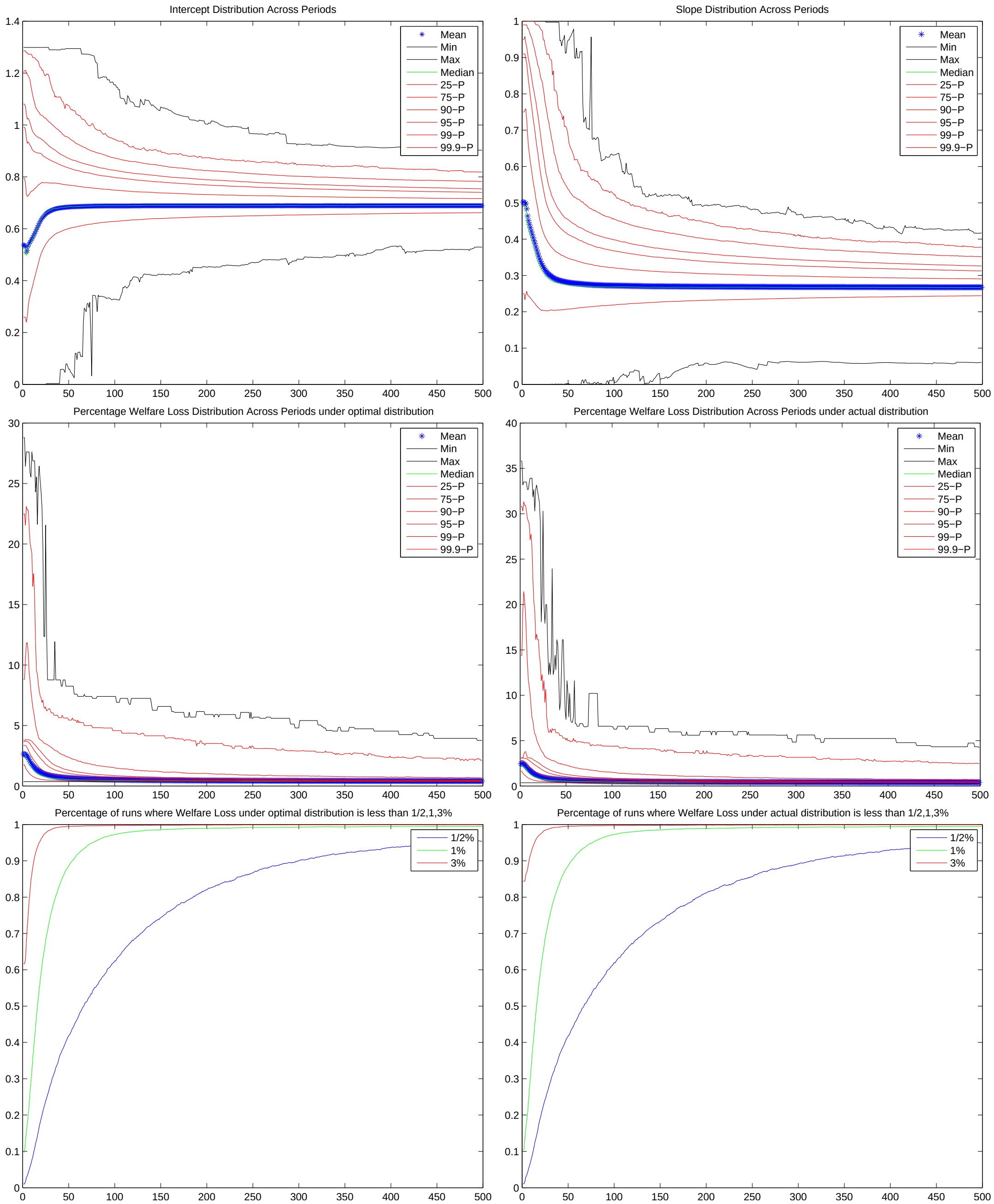

Figure 18: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 3.5$, $\delta = 0.9$.

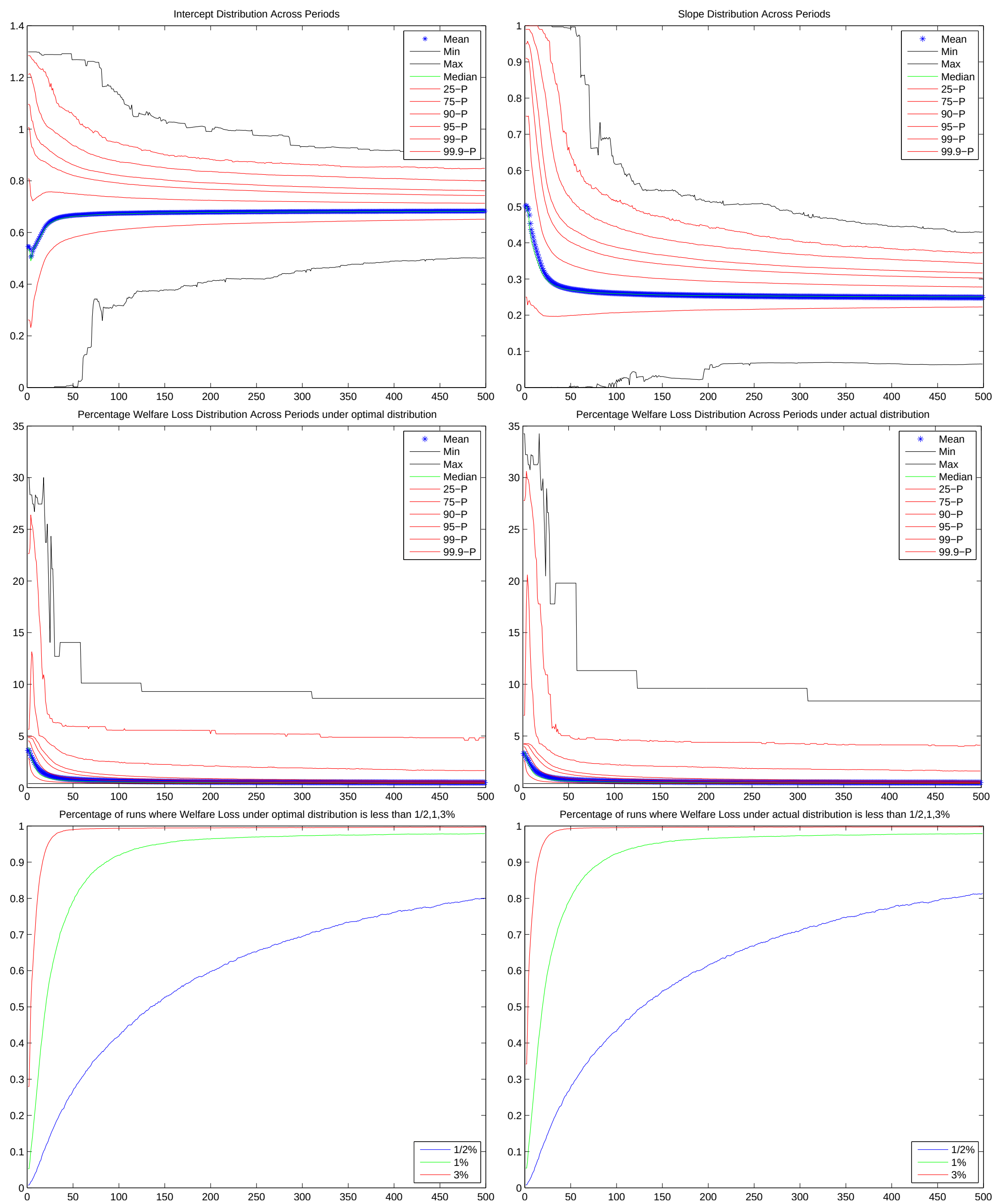

Figure 19: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 3.5$, $\delta = 0.95$.

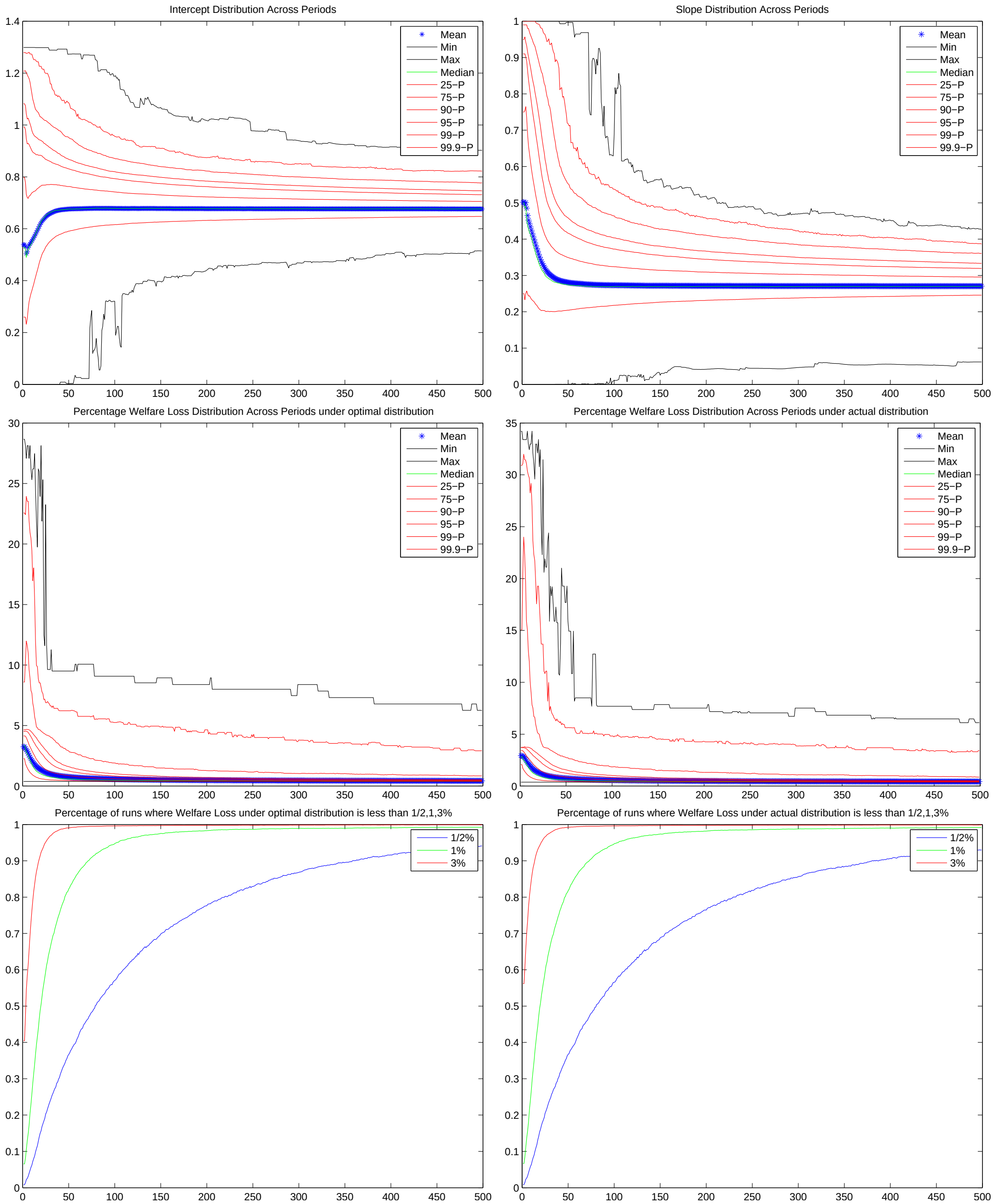

Figure 20: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 4$, $\delta = 0.9$.

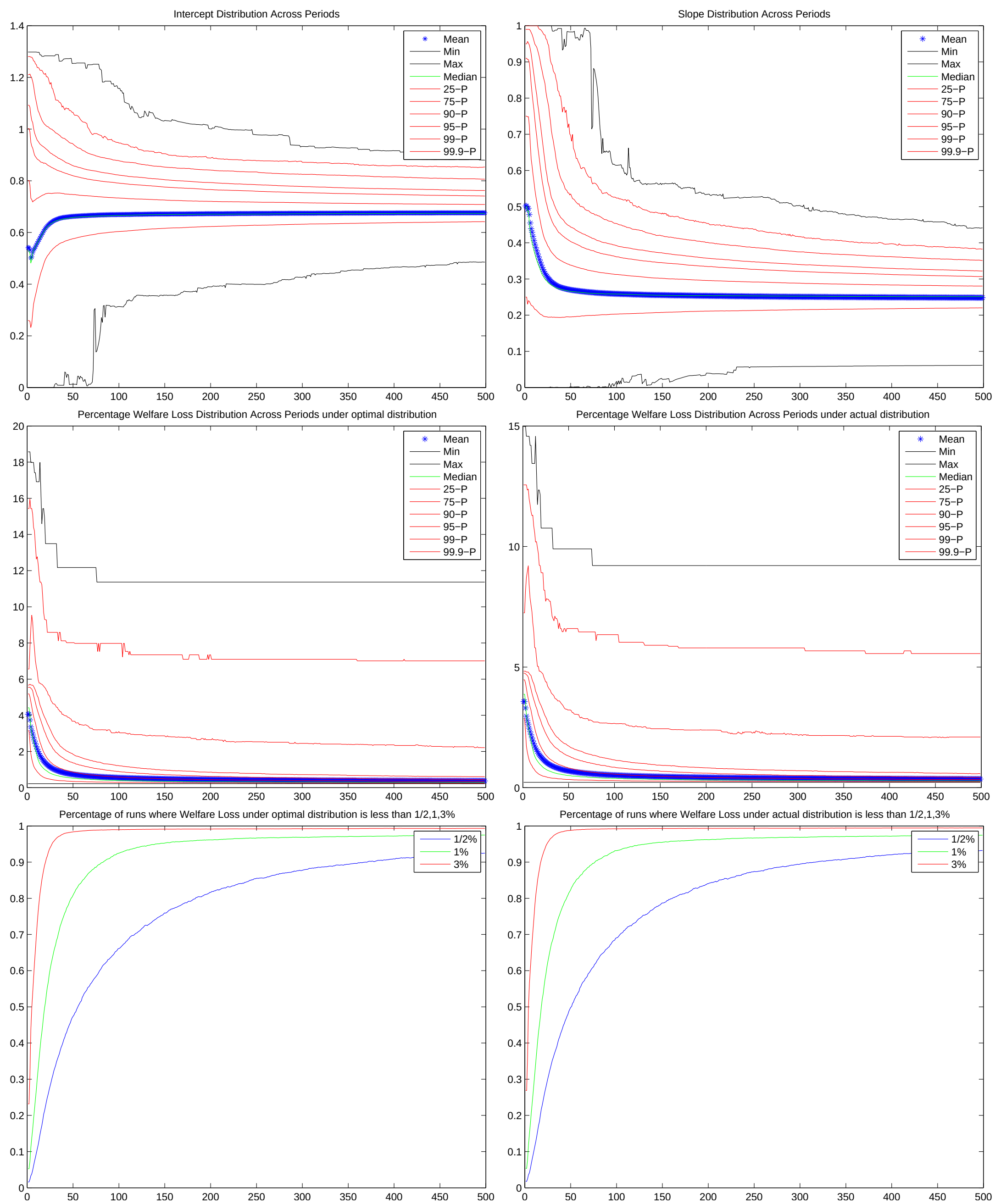

Figure 21: Distribution of α_t, β_t, $D_1^{\gamma_t}$, $D_2^{\gamma_t}$ and percentage of simulations with $D_1^{\gamma_t}$ and $D_2^{\gamma_t}$ less than $0.5\%, 1\%, 3\%$ for $\theta = 4$, $\delta = 0.95$.

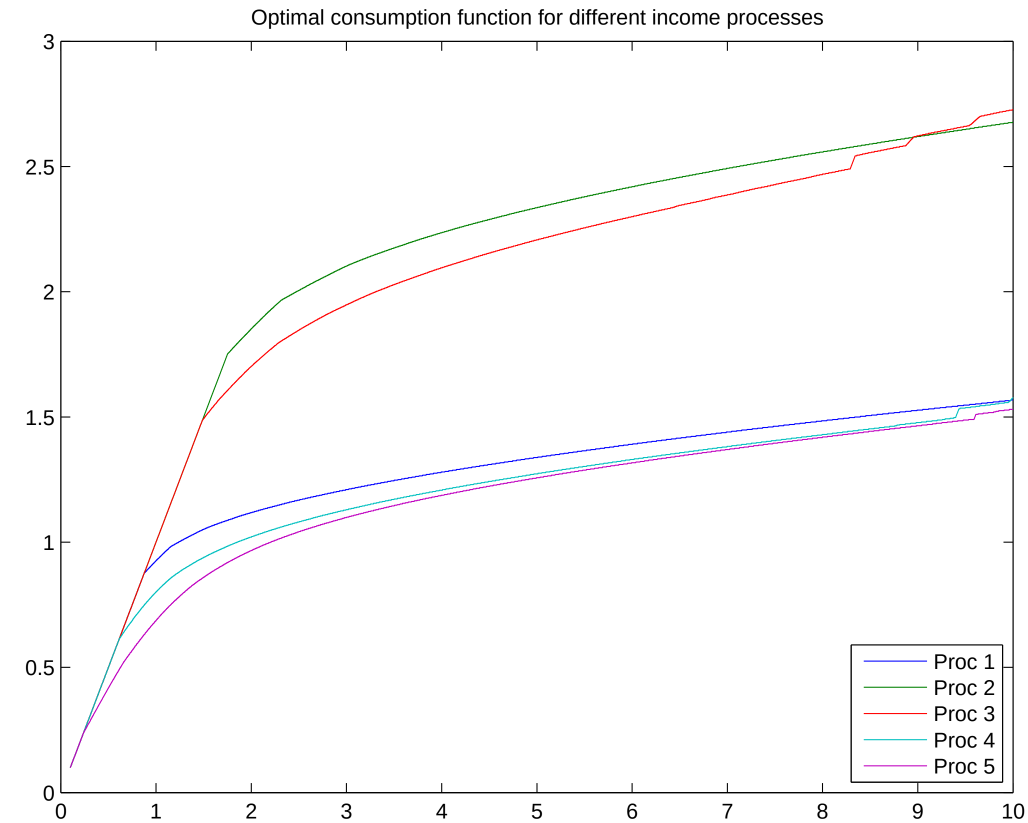

Figure 22: Optimal Consumption function for different income levels for a fixed parametrization with $\theta = 3.5$, $\delta = 0.95$, $B = 0$, $\eta = 1$, $\epsilon = 0$, $\xi = 1$ and $\phi = 1$.

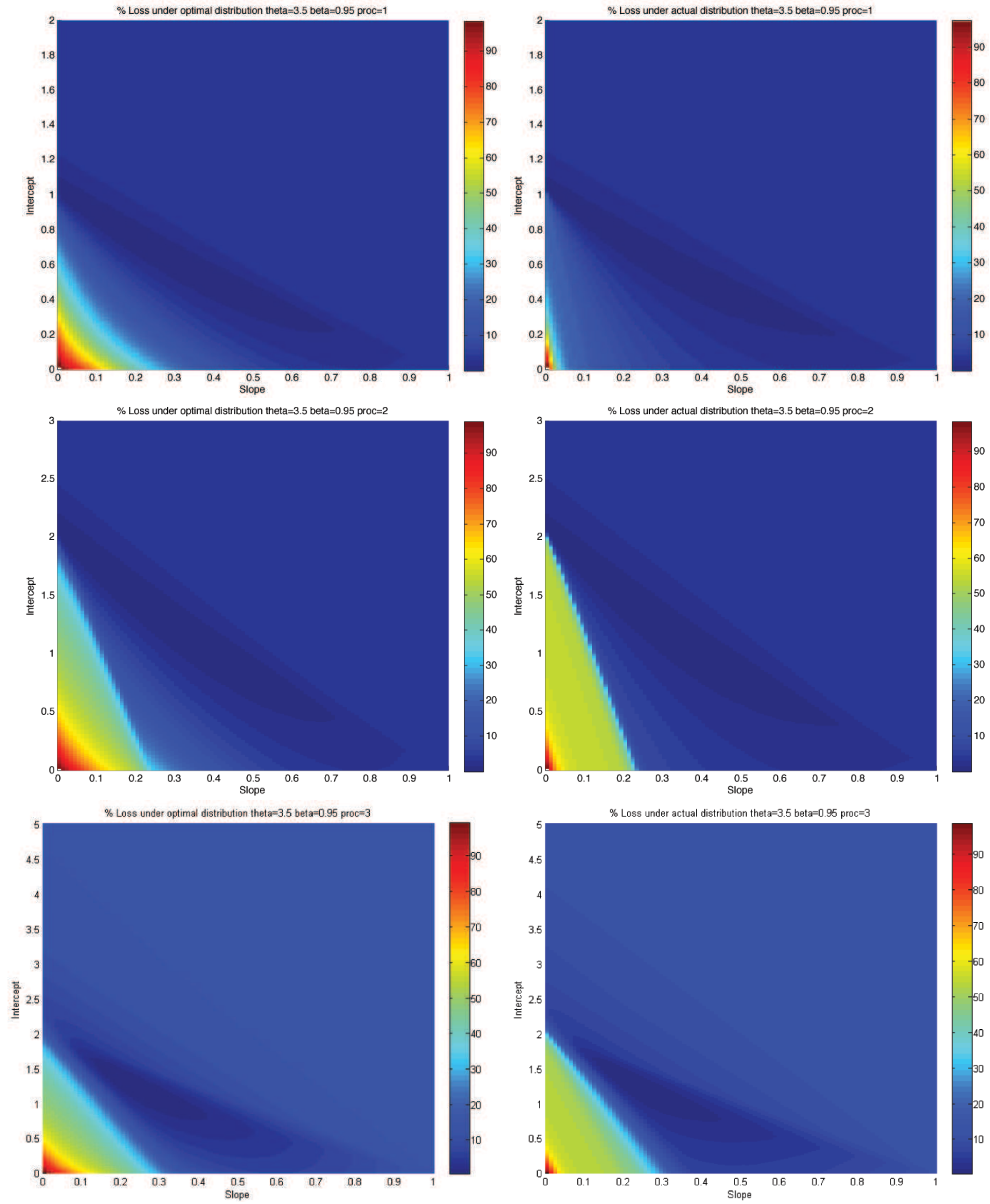

Figure 23: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept and slope, under π_* (left) and π_γ (right) for different income processes.

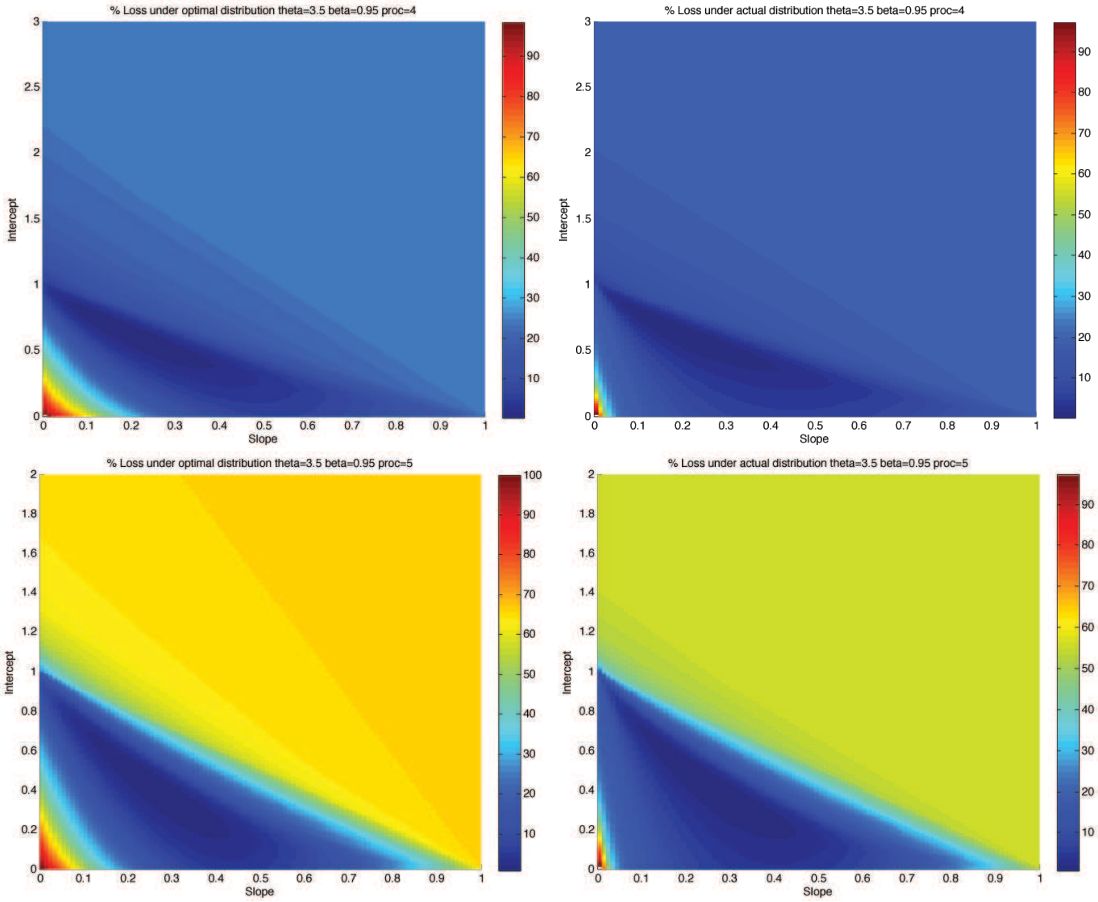

Figure 24: Percentage deviation of equivalent consumption of the linear rule from the optimal consumption function for different values of the intercept and slope, under π_* (left) and π_γ (right) for different income processes.

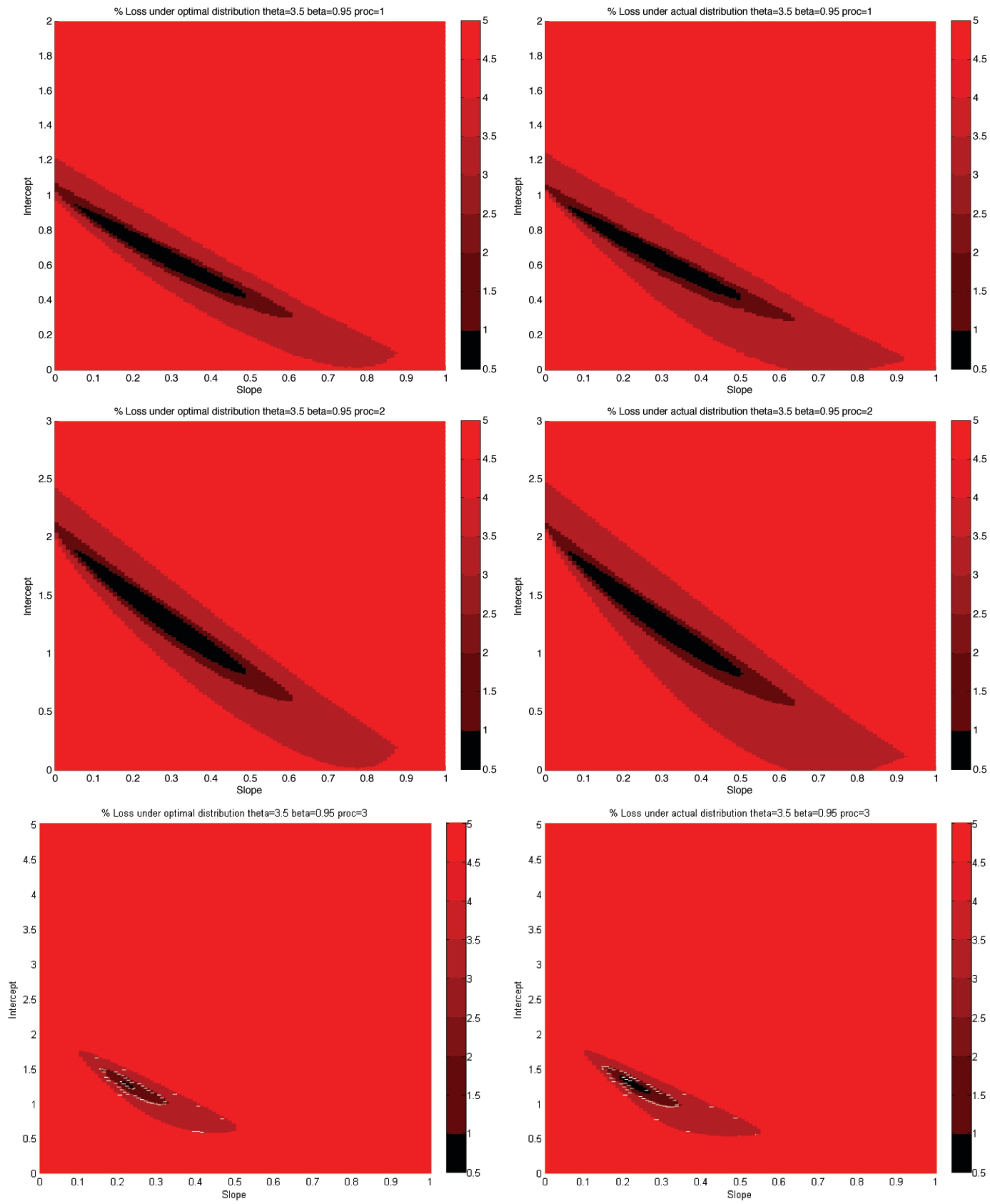

Figure 25: Set of consumption rules which have a percentage deviation less than or equal to 0.5% 1%, 3% and more than 3% for different income processes.

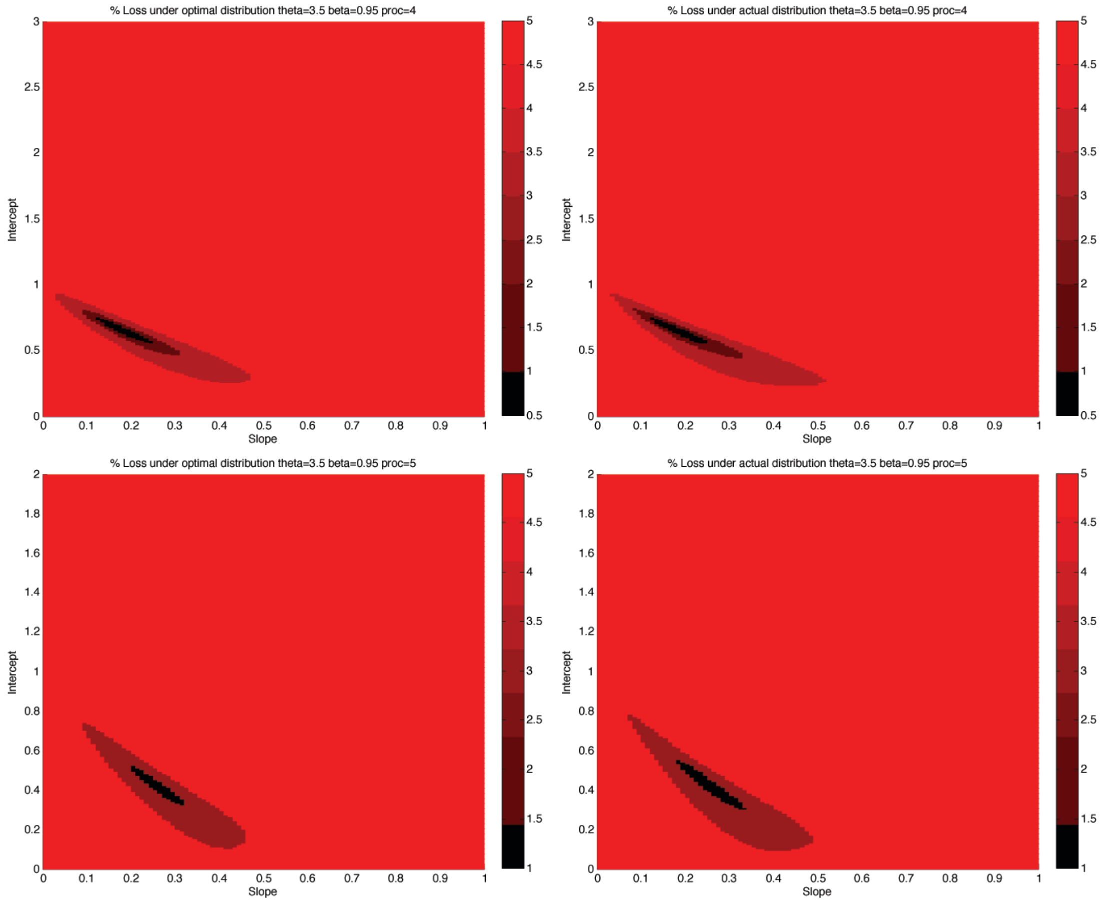

Figure 26: Set of consumption rules which have a percentage deviation less than or equal to 0.5% 1%, 3% and more than 3% for different income processes.

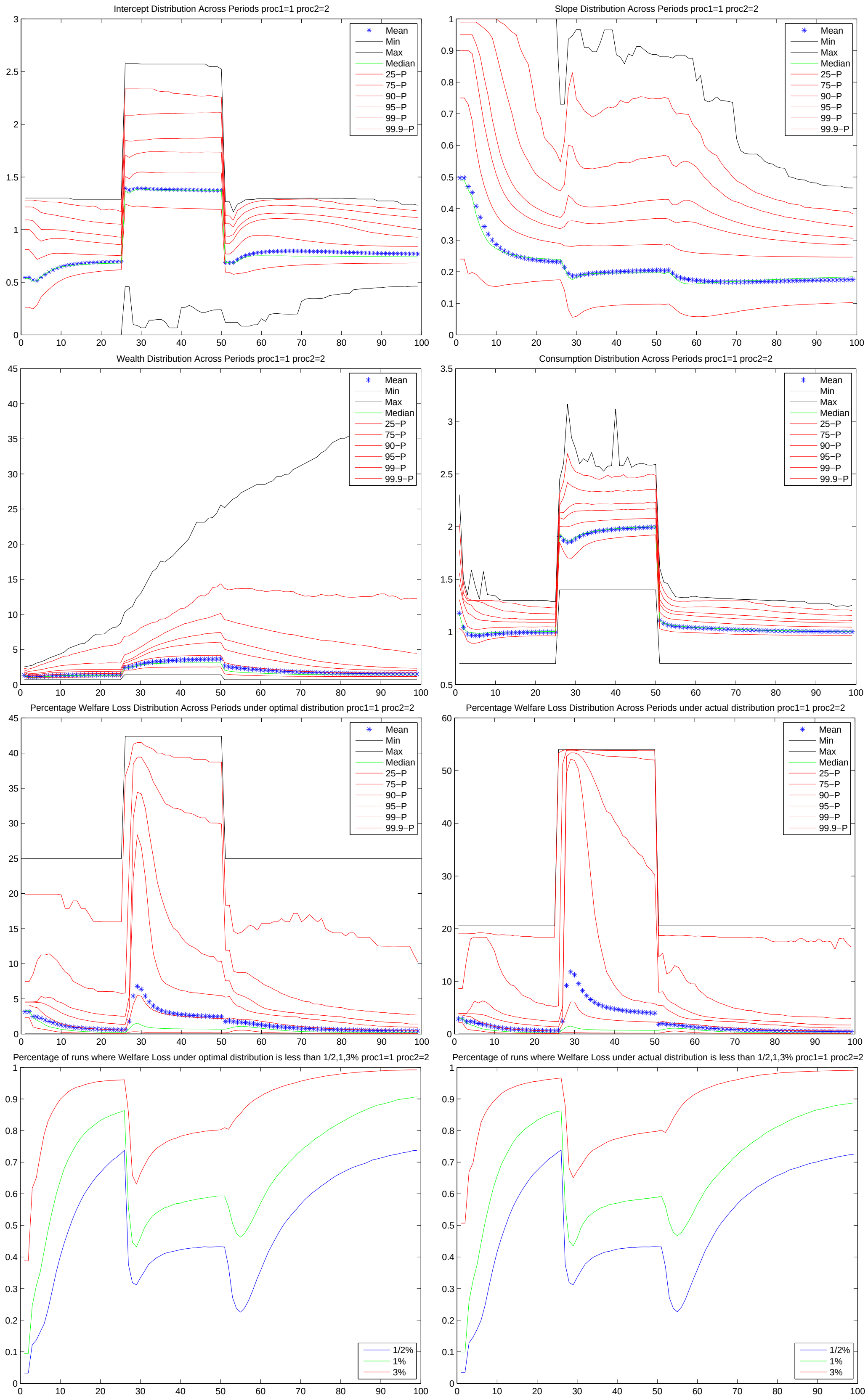

Figure 27: Dynamics under income shocks. During periods 1-25 and 51-100 agents' income follows Y^1. During periods 26-50 their income follows Y^2.

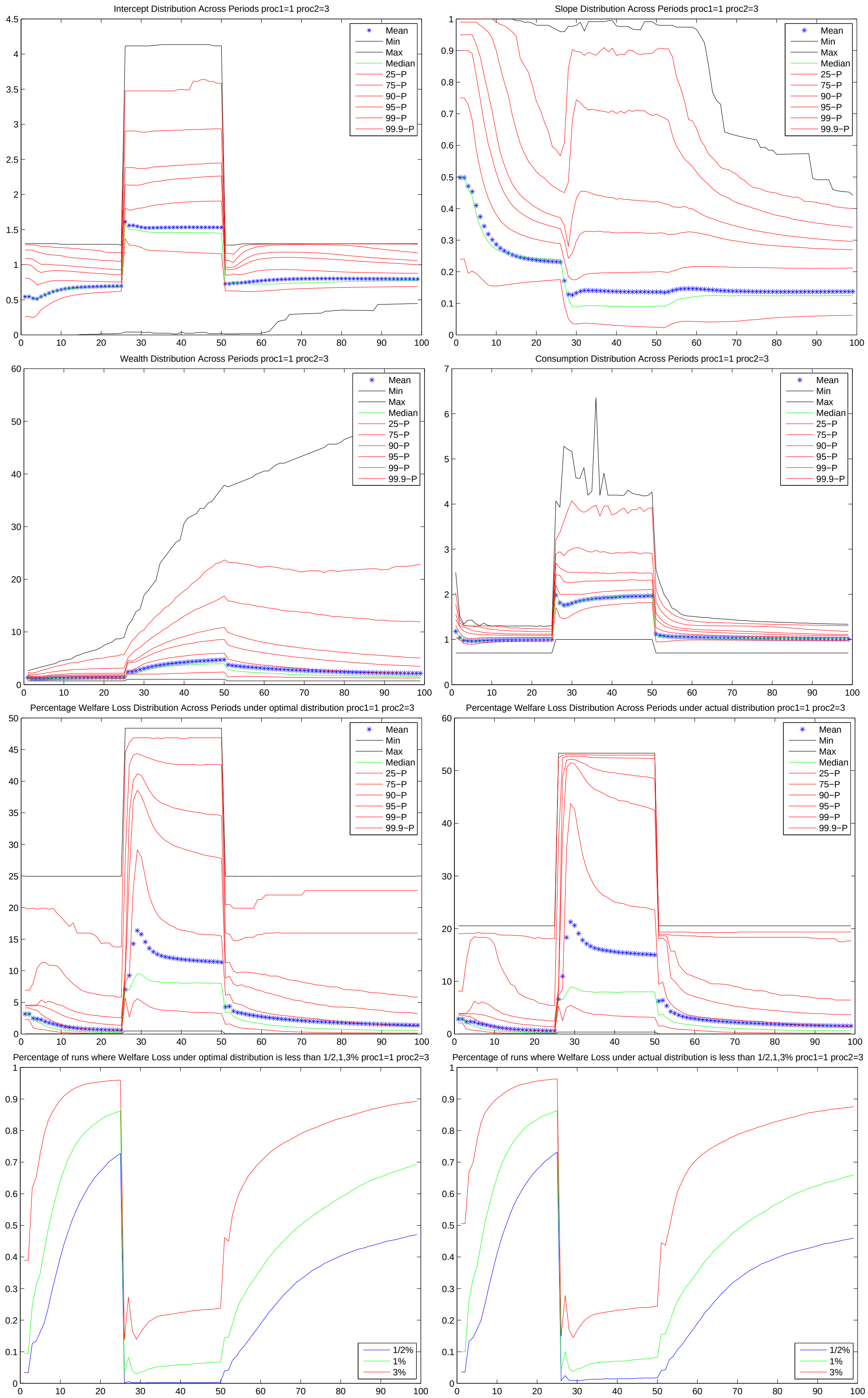

Figure 28: Dynamics under income shocks. During periods 1-25 and 51-100 agents' income follows Y^1. During periods 26-50 their income follows Y^3.

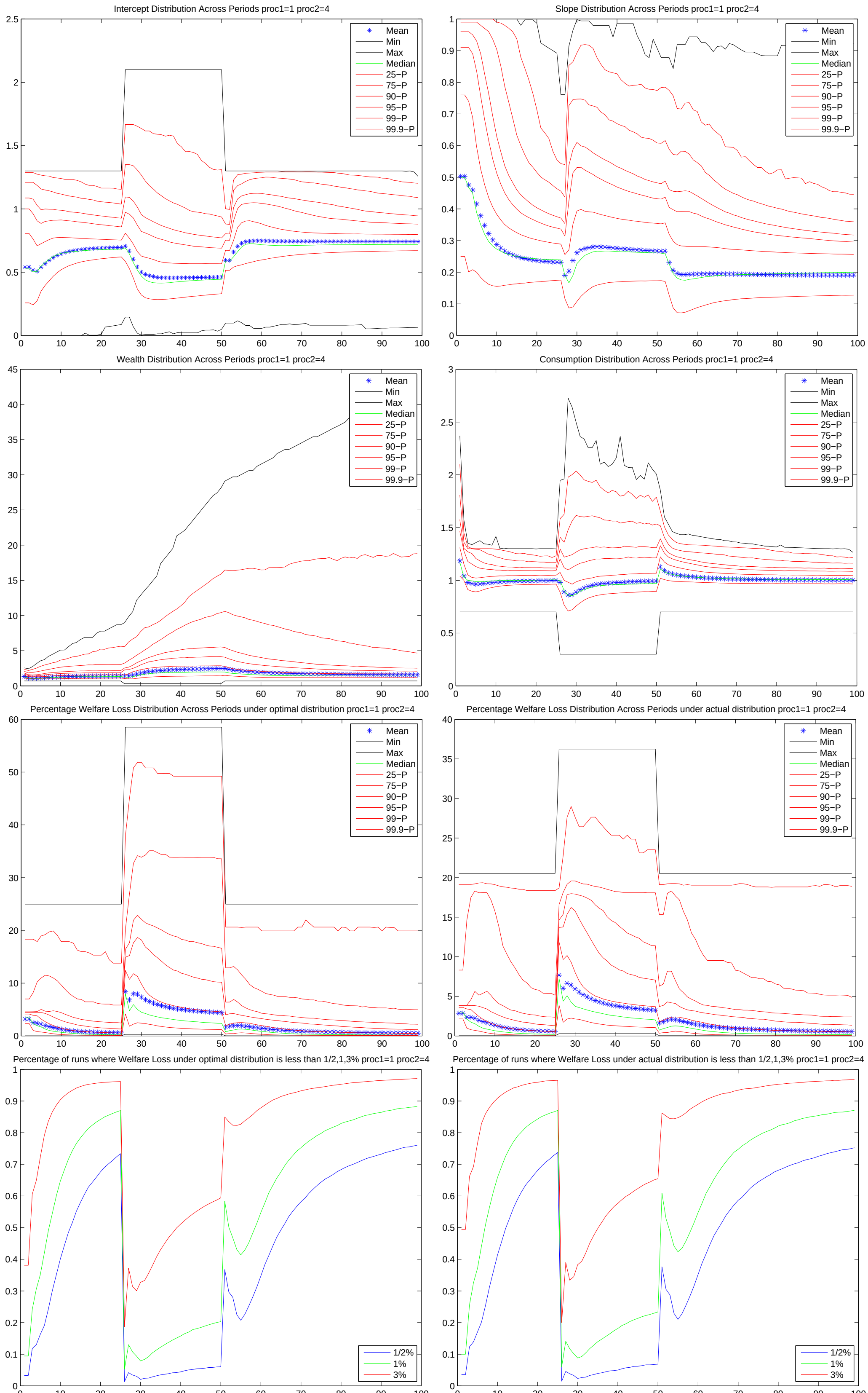

Figure 29: Dynamics under income shocks. During periods 1-25 and 51-100 agents' income follows Y^1. During periods 26-50 their income follows Y^4.

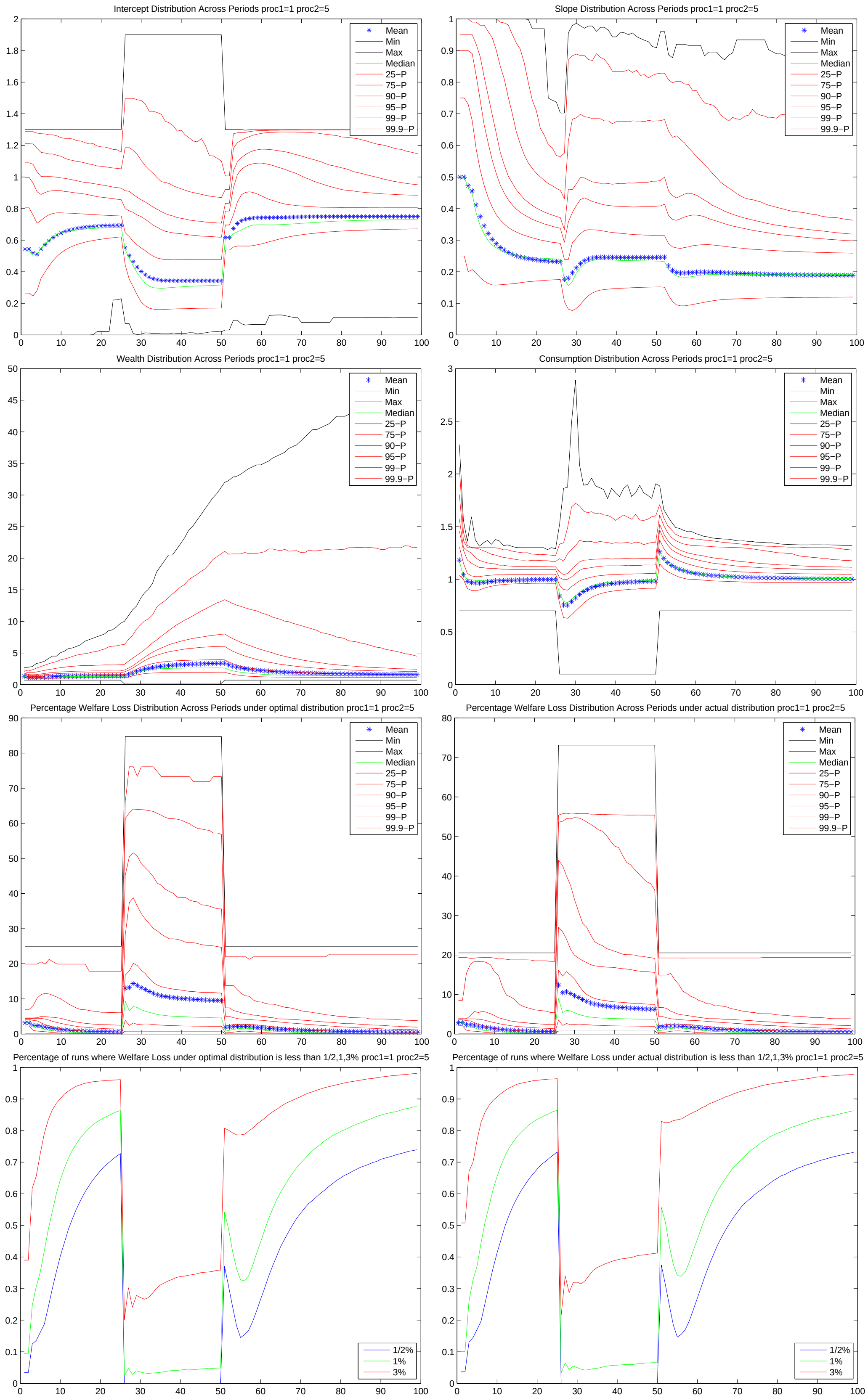

Figure 30: Dynamics under income shocks. During periods 1-25 and 51-100 agents' income follows Y^1. During periods 26-50 their income follows Y^5.

References

AKERLOF, G. A., AND J. L. YELLEN (1985a): "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?," *The American Economic Review*, 75(4), 708–720.

——— (1985b): "A Near-rational Model of the Business Cycle, with Wage and Price Intertia," *The Quarterly Journal of Economics*, 100(5), 823–38.

ALLEN, T. W., AND C. D. CARROLL (2001): "Individual Learning about Consumption," *Macroeconomic Dynamics*, 5(02), 255–271.

ARTHUR, W. (1994): "Inductive reasoning and bounded rationality," *American Economic Review*, 84(2), 406–411.

BROWN, A. L., Z. E. CHUA, AND C. F. CAMERER (2009): "Learning and Visceral Temptation in Dynamic Saving Experiments*," *Quarterly Journal of Economics*, 124(1), 197–231.

CARROLL, C. D. (2001): "A Theory of the Consumption Function, with and without Liquidity Constraints," *The Journal of Economic Perspectives*, 15(3), 23–45.

——— (2004): "Theoretical Foundations of Buffer Stock Saving," .

CARROLL, C. D., AND M. S. KIMBALL (1996): "On the Concavity of the Consumption Function," *Econometrica*, 64(4), 981–992.

CLARK, A. (1997): *Being there: putting brain, body, and world together again*. MIT Press, Cambridge, Mass.

COCHRANE, J. H. (1989): "The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives," *The American Economic Review*, 79(3), 319–337.

EVANS, G., AND B. MCGOUGH (2009): "Learning to Optimize," .

HOLLAND, J. H. (1992): *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.* MIT Press, Cambridge, Mass., 1st mit press ed edn.

JUDD, K. L. (1998): *Numerical methods in economics.* MIT Press, Cambridge, Mass.

KUSHNER, H. J., AND G. YIN (2003): *Stochastic approximation and recursive algorithms and applications*, vol. 35. Springer, New York.

LEIJONHUFVUD, A. (1993): "Towards a Not-Too-Rational Macroeconomics," *Southern Economic Journal*, 60(1), 1–13.

LETTAU, M., AND H. UHLIG (1999): "Rules of thumb versus dynamic programming," *American Economic Review*, 89(1), 148–174.

LJUNG, L. (1977): "Analysis of recursive stochastic algorithms," *IEEE transactions on automatic control*, 22(4), 551–575.

MARSHALL, A. (1890): *Principles of economics.* Macmillan and co., London and New York.

ÖZAK, Ö. (2009): "Optimal Consumption under Bounded Rationality and Liquidity Constraints," .

ROBBINS, H., AND S. MONRO (1951): "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400–407.

SATZ, D., AND J. FEREJOHN (1994): "Rational choice and social theory," *The Journal of philosophy*, 91(2), 71–87.

SIMON, H. (1990): "Invariants of Human Behavior," *Annual Reviews in Psychology*, 41(1), 1–20.

WATKINS, C. (1989): *Learning from delayed rewards.* Cambridge.

Weintraub, E. R. (1979): *Microfoundations: the compatibility of microeconomics and macroeconomics*, Cambridge surveys of economic literature. Cambridge University Press, Cambridge.